



Db 138 LGAWKNGANMAADAANKANAKDGTAGAIALTVMAGKGF 178

RESULT 3  
F83570  
hypothetical protein PA0599 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83570  
R:Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, N.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: F83570  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-553 <STO>  
A:Cross-references: GB:AE004496; GB:AE004091; NID:g9946468; PIDN:AAG03988.1; GSPDB:GN001  
C:Genetics:  
Experimental source: strain PA01  
A:Gene: PA0599

Query Match 26.8%; Score 63; DB 2; Length 353;  
Best Local Similarity 42.9%; Pred. No. 2.4;  
Matches 12; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

OY 1 PLTNPIDAIGSSADRNAEAFDKMKDD 28  
DB 137 PYTRPESASASADKADASADASKPD 164

RESULT 4  
F64202  
probable GTP-binding protein - Mycoplasma genitalium  
C:Species: Mycoplasma genitalium  
C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 02-Feb-2001  
C:Accession: F64202; T09682  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhmann, J.; Nguyen, D.; Uitterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A:Title: The minimal gene complement of Mycoplasma genitalium.  
A:Reference number: A64200; MUID:96026346  
A:Accession: F64202  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-367 <TIGR>  
A:Cross-references: GB:U39680; GB:L43967; NID:g1045681; PID:g1045693; TIGR:MG024  
A:Experimental source: strain G-37  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhmann, J.L.; Nguyen, D.T.; Uitterback, T.; Saudek, D.M.; Phillips, C.A.; Merrick, J.C.A.; Venter, J.C.  
submitted to the EMBL Data Library, October 1998  
A:Reference number: Z1818  
A:Accession: T09682  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-367 <FRA>  
A:Cross-references: EMBL:U39681; NID:g3844626; PID:g3844633  
C:Genetics:  
A:Gene: MG024  
A:Genetic code: SGC3  
C:Superfamily: yeast probable purine nucleotide-binding protein YBR025C  
C:Keywords: GTP; GTP binding; nucleotide binding; P-loop  
F:8-15/Region: nucleotide-binding motif A (P-loop)  
F:70-74/Region: GTP binding

Query Match 25.1%; Score 59; DB 2; Length 367;  
Best Local Similarity 34.8%; Pred. No. 8;

Matches 16; Conservative 9; Mismatches 15; Indels 6; Gaps 2;  
OY 2 LTNPIDAAIGSSADRNAEAFDKMKDDQIAAMVLRGMKDGCFAL 47  
DB 139 ITNRI-----GKLKRAESGDKIAKEEFLVEIVLNGL-KQCOMPI 178

RESULT 5  
T13432  
gamma-glutamyltransferase homolog T17A13.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 18-Feb-2000  
C:Accession: T13432  
R:Beyan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban  
submitted to the Protein Sequence Database, July 1999  
A:Reference number: Z17683  
A:Accession: T13432  
A:Molecule type: DNA  
A:Residues: 1-637 <BEV>  
A:Cross-references: EMBL:AL096692; GSPDB:GN00062; ATSP:T17A13.30  
A:Experimental source: cultivar Columbia; BAC clone F17A13  
C:Genetics:  
A:Gene: ATSP:T17A13.30  
A:Map position: 4  
A:Introns: 43/1; 156/3; 497/3; 543/3  
C:Superfamily: gamma-glutamyltransferase

Query Match 24.9%; Score 58.5; DB 2; Length 637;  
Best Local Similarity 28.1%; Pred. No. 16;  
Matches 18; Conservative 8; Mismatches 19; Indels 19; Gaps 2;

OY 2 LTNPIDAAIGSSA-----DRNAEAFD-----KKMKDDQIAAMVLRGMKAD 42  
DB 117 VYVPMSSSGIGSGSFLVSSQKSKAEAFDMRTAPLASKDKMKKADSAKSIGALSMGVP 176

OY 43 GQFA 46  
DB 177 GR1A 180

RESULT 6  
H86191  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86191  
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizlar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Morzila  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: H86191  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-500 <STO>  
A:Cross-references: GB:AE005172; NID:g6850304; PIDN:AAF29381.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 24.7%; Score 58; DB 2; Length 500;  
Best Local Similarity 37.5%; Pred. No. 15;  
Matches 12; Conservative 5; Mismatches 15; Indels 0; Gaps 0;  
OY 12 GSADRNAEAFDKMKDDQIAAMVLRGMKDG 43  
DB 139 ITNRI-----GKLKRAESGDKIAKEEFLVEIVLNGL-KQCOMPI 178

Db 154 GOVDNAKMPDKMERDLISWTAMINGFVKKG 185

## RESULT 7

Variable major protein 21 - Borrelia hermsli

C:Species: Borrelia hermsli

C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999

C:Accession: S11981

R:Burman, N.; Bergstroem, S.; Restrepo, B.I.; Barbour, A.G.

Mol. Microbiol. 4, 1715-1726, 1990

A:Title: The variable antigens Vmp7 and Vmp21 of the relapsing fever bacterium Borrelia

A:Reference number: S11980; MUID:91171872

A:Accession: S11981

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <BMR>

A:Cross-references: GB:M57256; EMBL:X53927; NID:G144025; PIDN:AA859031.1; PID:G144026

## Query Match

23.8%; Score 56; DB 2; Length 364;

Best Local Similarity 29.8%; Pred. No. 19;

Matches 14; Conservative 10; Mismatches 15; Indels 8; Gaps 2;

3 TNPIDAIGSA---DNAEAFDMMKKDDQIAAAMVLRGMKDGQFA 46

Db 267 TTPLEFAVGNGCAHLSONANS---KAAVAAGTALRLSVKGGKLA 308

## RESULT 8

hypothetical protein AAF81289.1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: D86269

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federle, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.R.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: D86269

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-797 <STO>

A:Cross-references: GB:AE05172; NID:98920567; PIDN:AAF81289.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

## Query Match

23.8%; Score 56; DB 2; Length 797;

Best Local Similarity 34.2%; Pred. No. 42;

Matches 13; Conservative 8; Mismatches 13; Indels 4; Gaps 1;

12 GSADRNAEAFDMMKKD---DOIAAAMVLRGMKDGQF 45

Db 390 GRIDEALSLFNOMKADGSLPLVAVSIYHGLCKLGRF 427

## RESULT 9

559069

213 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 28-May-1999

C:Accession: S59069

R:Schulz, T.C.; Hopwood, B.; Rathjen, P.D.; Wells, J.R.E.

Biochem. J. 311, 219-224, 1995

A:Title: An unusual arrangement of 13 zinc fingers in the vertebrate gene 213.

A:Reference number: S59069; MUID:96003919

A:Accession: S59069

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-794 <SCH>

A:Cross-references: GB:U14556; NID:9608136; PIDN:AAA8493.1; PID:9608137

C:Superfamily: POZ domain homology

F:10-108/Domain: POZ domain homology <POZ>

## Query Match

23.2%; Score 54.5; DB 2; Length 794;

Best Local Similarity 42.9%; Pred. No. 64;

Matches 18; Conservative 6; Mismatches 15; Indels 3; Gaps 3;

2 LTNPIDAIGSADRNA-EAFDMMKKDDQIAAAMVLRGMK 41

Db 114 LAEP-SFTGESADASAVEGDKRAKDEKAATMLSRGQAR 154

## RESULT 10

hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: Columbia

C>Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998

C:Accession: F71401

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D.

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; G

avanagh, T.; Hempel, S.; Kotter, P.; Ertlian, K.D.; Kieger, M.; Schaeffer, M.; Funk, B

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen

erhoff, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113

A:Accession: F71401

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-258 <BEV>

A:Cross-references: GB:297335; NID:G2244747; PID:E326865; PID:G2244761

C:Genetics:

A:Map position: 4COP9-4G3845

## Query Match

23.0%; Score 54; DB 2; Length 258;

Best Local Similarity 36.6%; Pred. No. 23;

Matches 15; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

1 PLTNPIDAIG--GSADRNAEAFDMMKKDDQIAAAMVLRGM 39

Db 107 PLANTLVNYYGKCAASHALQVEDPMPHRDITAMASVLTAL 147

## RESULT 11

744962

succinate dehydrogenase chain A homolog [imported] - Natronomonas pharaonis

C:Species: Natronomonas pharaonis

C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Mar-2000

C:Accession: T44962

R:Matlar, S.; Souquet, M.; Henrich, H.J.; Engelhard, M.

submitted to the EMBL Data Library, August 1996

A:Description: The first fully resolved primary structure of an archaeal succinate-de

A:Reference number: 222881

A:Accession: T44962

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-611 <MAT>

A:Cross-references: EMBL:Y07709; PIDN:CAA68982.1

A:Experimental source: strain SP1 /28

C:Genetics:

A:Gene: sdhA

C:Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology

Query Match 23.0%; Score 54; DB 2; Length 611;  
Best Local Similarity 35.1%; Pred. No. 57;  
Matches 13; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

OY 6 IDAIGSADRNAEAFDKMKDDQIAAAMVLRGMKD 42  
DB 51 INAAIRGDDWELHAYDPMKGSYLGDAIPAETTLAOD 87

## RESULT 12

B85153  
hypothetical protein AT4g14050 (imported) - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: B85153  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Accession: B85153  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-686 <STO>  
A:Cross-references: GB:NC\_001268; NID:97268109; PIDN:CAB78447.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g14050  
A:Map position: 4  
C:Superfamily: Arabidopsis thaliana hypothetical protein At2g41080

Query Match 23.0%; Score 54; DB 2; Length 686;  
Best Local Similarity 36.6%; Pred. No. 64;  
Matches 15; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

OY 1 PUTNPIDAAIG--GSADRNAEAFDKMKDDQIAAAMVLRGM 39  
DB 107 PLANTLVNYYGCGAASHALQVDEMPHRDHAMASVLTAL 147

## RESULT 13

T33030  
hypothetical protein M70.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T33030  
R:Miller, N.; Wamsley, P.; Twyman, B.  
A:Submitted to the EMBL Data Library, February 1998  
A:Description: The sequence of C. elegans cosmid M70.  
A:Reference number: Z21266  
A:Accession: T33030  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-915 <ML>  
A:Cross-references: EMBL:AF047661; PIDN:AAC04437.1; GSPDB:GN00022; CESP:M70.3  
A:Experimental source: strain Bristol N2; clone M70  
C:Genetics:  
A:Gene: CESP:M70.3  
A:Map position: 4  
A:Introns: 20/1; 49/1; 78/1; 98/1; 793/2; 891/3

Query Match 23.0%; Score 54; DB 2; Length 915;  
Best Local Similarity 45.5%; Pred. No. 86;  
Matches 15; Conservative 2; Mismatches 10; Indels 6; Gaps 1;

OY 5 PIDAA-----IGSADRNAEAFDKMKDDQIA 31  
DB 130 PSDEAITDLNIGGVKLNKDLFFDKTKVDVIA 162

## RESULT 14

T33032  
hypothetical protein M70.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T33032  
R:Miller, N.; Wamsley, P.; Twyman, B.  
A:Submitted to the EMBL Data Library, February 1998  
A:Description: The sequence of C. elegans cosmid M70.  
A:Reference number: Z21266  
A:Accession: T33032  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-941 <ML>  
A:Cross-references: EMBL:AF047661; PIDN:AAC04435.1; GSPDB:GN00022; CESP:M70.1  
A:Experimental source: strain Bristol N2; clone M70  
C:Genetics:  
A:Gene: CESP:M70.1  
A:Map position: 4  
A:Introns: 20/1; 40/1; 60/1; 80/1; 774/2; 892/3

Query Match 23.0%; Score 54; DB 2; Length 941;  
Best Local Similarity 45.5%; Pred. No. 88;  
Matches 15; Conservative 2; Mismatches 10; Indels 6; Gaps 1;

OY 5 PIDAA-----IGSADRNAEAFDKMKDDQIA 31  
DB 112 PSDEAITDLNIGGVKLNKDLFFDKTKVDVIA 144

## RESULT 15

T25888  
hypothetical protein T10H10.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000  
C:Accession: T25888  
R:Nelson, J.; Langston, Y.  
A:Submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid T10H10.  
A:Reference number: Z20105  
A:Accession: T25888  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2098 <NEL>  
A:Cross-references: EMBL:U80848; PIDN:AAB37988.1; GSPDB:GN00028; CESP:T10H10.1  
A:Experimental source: strain Bristol N2; clone T10H10  
C:Genetics:  
A:Gene: CESP:T10H10.1  
A:Map position: X  
A:Introns: 6/3; 43/3; 92/3; 154/2; 226/1; 280/3; 417/2; 472/3; 812/3; 853/3; 887/3; 9/3  
C:Superfamily: myosin motor domain homology  
F:65-720/Domain: myosin motor domain homology <MCO>

Query Match 22.8%; Score 53.5; DB 2; Length 2088;  
Best Local Similarity 48.1%; Pred. No. 2.3e+02;  
Matches 13; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

OY 21 FDKMKDDQI-AAAMVLRGMKDGQFA 46  
DB 1903 YHKTKNVIELAALILRSMTKDKKNA 1929

Search completed: January 10, 2002, 14:04:30  
Job time: 124 sec



```

RESULT 2
ID Y024_MYCGE STANDARD: PRT: 367 AA.
AC P47270:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GTP-BINDING PROTEIN MG024.
CM MG024.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
R MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.W.,
RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lacer T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RT Science 270:397-403(1995).
CC -1 SIMILARITY: BELONGS TO THE YOHF FAMILY OF GTP-BINDING PROTEINS
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U39681; AAC71240.1; -.
DR TIGR: MG024; -.
DR InterPro: IPR000765; GTP1_OBG.
DR Pfam: PF01018; GTP1_OBG.1.
DR PRINTS: PR00326; GTP1_OBG.
KM GTP-binding; Complete proteome.
FT NP_BIND 8 15 GTP (POTENTIAL).
FT NP_BIND 71 75 GTP (POTENTIAL).
FT FT
SQ SEQUENCE 367 AA: 41198 MM; 98AFCFADE6C9BDC CRC64;

```

```

Query Match 25.1%; Score 59; DB 1; Length 367;
Best Local Similarity 34.8%; Pred. No. 3.6;
Matches 16; Conservative 9; Mismatches 15; Indels 6; Gaps 2;

```

```

OY 2 LTPIDAIGSADRNAEFDKKKDDQIAAMVLRGMKADGQFA 47
DB 139 ITNR1-----GKLRKASGDKIAKEEVLLEIVLNGL-KQGGMP 178

```

```

RESULT 3
ID VM21_BORHE STANDARD: PRT: 364 AA.
AC P21875;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 21 PRECURSOR.
CM VM21.
OS Borrelia hermsli.
OS Plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HSI SEROTYPE 21;

```

```

RX MEDLINE=91171872; PubMed=1706456;
RA Burman N., Bergstrom S., Restrepo B.I., Barbour A.G.;
RT "The variable antigens Vmp7 and Vmp21 of the relapsing fever
RT bacterium Borrelia hermsli are structurally analogous to the VSG
RT proteins of the African trypanosome.";
RL Mol. Microbiol. 4:1715-1726(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HSI SEROTYPE 21;
R MEDLINE=93133110; PubMed=1484486;
RA Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT "Subtelomeric expression regions of Borrelia hermsli linear plasmids
RT are highly polymorphic.";
RL Mol. Microbiol. 6:3299-3311(1992).
RN [3]
RP SEQUENCE OF 55-75; 186-208 AND 245-259.
RX MEDLINE=85236116; PubMed=2409197;
RA Barsiad P.A., Colligan J.E., Raum M.G., Barbour A.G.;
RT "Variable major proteins of Borrelia hermsli. Epitope mapping and
RT partial sequence analysis of CNBR peptides.";
RL J. Exp. Med. 161:1302-1314(1985).
CC -1 FUSION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC -1 SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC -1 SIMILARITY: STRONG, TO VMP7 AND VMP25.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M57256; AAB59031.1; -.
DR PIR: S11981; S11981.
DR InterPro: IPR000680; Borrelia_lipo.
DR Pfam: PF00921; Lipoprotein_2.1.
DR PROSITE: PS0013; PROKAR_LIPOPROTEIN_1.
KM Outer membrane; Lipoprotein; Signal; Plasmid.
FT SIGNAL 1 26 PROBABLE.
FT CHAIN 27 364 VARIABLE MAJOR OUTER MEMBRANE
FT FT LIPOPROTEIN 21.
FT FT N-ACYL DIGLYCERIDE (PROBABLE).
FT FT
SQ SEQUENCE 364 AA: 37084 MM; 16598B635EB3776 CRC64;

```

```

Query Match 23.8%; Score 56; DB 1; Length 364;
Best Local Similarity 29.8%; Pred. No. 8.5;
Matches 14; Conservative 10; Mismatches 15; Indels 8; Gaps 2;

```

```

OY 3 TNPIDAIGGSA---DRNAEAFDKKKDDQIAAMVLRGMKADGQFA 46
DB 267 TTPLFAVAGNGAHLNOMANS-----KAAVAVGGIAIRSLVKGRLA 308

```

```

RESULT 4
ID Z151_MOUSE STANDARD: PRT: 794 AA.
AC Q60821; Q60699;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN 151 (POLYOMAVIRUS LATE INITIATOR PROMOTER BINDING
DE PROTEIN) (LP-1) (ZINC FINGER PROTEIN Z15).
CM ZNF151 OR ZFP100.
GN ZNF151 OR ZFP100.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Rapp L., Carmichael G.C.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBA; TISSUE=Kidney;  
RX MEDLINE=96003919; PubMed=7575457;  
RA Schulz T.C., Hopwood B., Rathjen P.D., Wells J.R.E.;  
RT "An unusual arrangement of 13 zinc fingers in the vertebrate gene  
Z13.";  
RL Biochem. J. 311:219-224(1995).  
CC -1- FUNCTION: MAY FUNCTION AS A HOUSEKEEPING DNA-BINDING PROTEIN THAT  
CC REGULATES THE EXPRESSION OF SPECIFIC GENES.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- TISSUE SPECIFICITY: FOUND IN ALL THE EMBRYONIC AND ADULT TISSUES  
CC EXAMINED.  
CC -1- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -1- SIMILARITY: CONTAINS 1 BTF/POZ DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U22396; AAA64848.1; -;  
DR EMBL: U14556; AAA85493.1; -;  
DR HSSP: P08047; ISP2.  
DR MGD: MGI:107410; Zfp100.  
DR InterPro: IPR000210; Zfp100.  
DR InterPro: IPR000822; Znf-C2H2.  
DR Pfam: PF00651; BTF; 1.  
DR Pfam: PF00096; Zf-C2H2; 13.  
DR PRINTS: PR00048; ZINCFINGER.  
DR SMART: SM00225; BTF; 1.  
DR SMART: SM00355; Znf-C2H2; 13.  
DR PROSITE: PS50097; BTF; 1.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 13.  
DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 13.  
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
KW Nuclear protein; Repeat.  
FT DOMAIN 1 104 BTF.  
FT DOMAIN 297 730 ZINC\_FINGERS.  
FT ZN\_FING 297 319 C2H2-TYPE.  
FT ZN\_FING 325 347 C2H2-TYPE.  
FT ZN\_FING 353 375 C2H2-TYPE.  
FT ZN\_FING 381 403 C2H2-TYPE.  
FT ZN\_FING 409 431 C2H2-TYPE.  
FT ZN\_FING 437 459 C2H2-TYPE.  
FT ZN\_FING 465 487 C2H2-TYPE.  
FT ZN\_FING 493 515 C2H2-TYPE.  
FT ZN\_FING 519 543 C2H2-TYPE.  
FT ZN\_FING 549 571 C2H2-TYPE.  
FT ZN\_FING 577 599 C2H2-TYPE.  
FT ZN\_FING 605 628 C2H2-TYPE.  
FT ZN\_FING 708 730 C2H2-TYPE.  
FT CONFLICT 507 507 G -> A (IN REF. 2).  
FT CONFLICT 573 573 N -> K (IN REF. 2).  
SQ SEQUENCE 794 AA; 86664 MW; FFF8856DEBDF7ED CAC64;

Query Match 23.2%; Score 54.5; DB 1; Length 794;  
Best Local Similarity 42.9%; Pred. No. 30;  
Matches 18; Conservative 6; Mismatches 15; Indels 3; Gaps 3;

QY 2 LTNPIDAIGSADRNA-EAFDKMKKKDOIAAAMLR-GMAK 41  
DB 114 LAEP-SSTGESADSAVSGDGKRAKDEKAATMTLSRLGQAR 154

RESULT 5

STCD\_RHIME  
ID STCD\_RHIME STANDARD: PRT: 580 AA.  
AC 087278;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE N-METHYLPROLINE DEMETHYLASE (EC 1.-.-.-) (STACHYDRINE  
DE UTILIZATION PROTEIN STCD).  
GN STCD.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OG Plasmid pSymB (megaplasmid 2).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RC STRAIN=1021;  
RX MEDLINE=98432824; PubMed=9758825;  
RA Phillips D.A., Sande E.S., Vriezen J.A.C., de Bruijn F.J.,  
RA Le Rudulier D., Joseph C.M.;  
RT "A new genetic locus in Sinorhizobium meliloti is involved in  
RT stachydrine utilization.";  
RL Appl. Environ. Microbiol. 64:3954-3960(1998).  
CC -1- FUNCTION: POSSIBLE NADH-DEPENDENT OXIDASE, FUNCTIONS AS A  
CC DEMETHYLASE THAT CONVERTS N-METHYLPROLINE TO PROLINE.  
CC -1- COFACTOR: FAD (POTENTIAL).  
CC -1- PATHWAY: SECOND STEP IN STACHYDRINE DEGRADATION.  
CC -1- INDUCTION: BY STACHYDRINE.  
CC -1- SIMILARITY: TO THE OYE FAMILY OF NADH-DEPENDENT FLAVIN  
CC OXIDOREDUCTASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF016307; AAC63907.1; -;  
DR InterPro: IPR000205; NAD-binding.  
DR InterPro: IPR001155; Oxidored.FMN.  
DR Pfam: PF00724; Oxidored.FMN; 1.  
KW Oxidoreductase; NAD; FAD; Flavoprotein; Plasmid.  
SQ SEQUENCE 580 AA; 63938 MW; 40BDAD5C5F096F3D CAC64;

Query Match 22.8%; Score 53.5; DB 1; Length 580;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

QY 1 PLTNPIDAIGSAD-RNAEAFDKMK 25  
DB 179 PLTNELDGYGSLNRRMRCFDVLK 204

RESULT 6  
YIS4\_YEAST  
ID YIS4\_YEAST STANDARD: PRT: 432 AA.  
AC P40564;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 48.6 KDA PROTEIN IN BET1-PANI INTERGENIC REGION.  
GN YIR004W OR YIB4W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,

RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
 RA Louis E., Lye G., Moule S., Nodell C., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.,  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RC MEDLINE-95282515; PubMed-7762303;  
 RA Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,  
 RA Schwager C., Zimmermann J., Sander C., Anstorge W.,  
 RT Nucleotide sequence and analysis of the centromeric region of yeast  
 RT chromosome IX.  
 RL Yeast 11:61-78(1995).  
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z38062; CAAB6206.1;  
 DR EMBL: X79743; NOT ANNOTATED\_CDS.  
 DR PIR: S48438; S48438.  
 DR HSSP: P08622; 1XBL.  
 DR SGP: S0001443; YIR004W.  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR Pfam: PF00226; DnaJ\_1.  
 DR SMART: SM00271; DnaJ\_1.  
 DR PROSITE: PS00636; DnaJ\_1;  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 DR Hypothetical protein; Chapetone.  
 KW DOMAIN  
 FT DOMAIN 4 73 J-DOMAIN  
 SQ SEQUENCE 432 AA; 48574 MW; 8127D686BC78B96 CRC64;

Query Match 22.3%; Score 52.5; DB 1; Length 432;  
 Best Local Similarity 29.3%; Pred. No. 28;  
 Matches 12; Conservative 8; Mismatches 10; Indels 11; Gaps 1;

OY 3 TNPDAIGSGADR-----NAEAFDKMKKDDQIAA 32  
 DB 145 TAAVDAAGNTNEKDKKARTTSGULTVHDGKKNQYGA 185

RESULT 7  
 ID SR54\_HALN1 STANDARD; PRT; 460 AA.  
 AC O9HNS:  
 DT 20-AUG-2001 (rel. 40, Created)  
 DT 20-AUG-2001 (rel. 40, Last sequence update)  
 DE SIGNAL RECOGNITION 54 KDA PROTEIN (SRP54).  
 GN SRP54 OR VNG2459G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 CC Halobacterium.  
 NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-20504483; PubMed-11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasfy S.R., Baliga N.S., Thorsson V., Shrogha J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Meli R., Goo Y.A.,  
 RA Leitner B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddock D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Beck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;

RT "Genome sequence of Halobacterium species NRC-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -1- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECRETORY PROTEIN  
 CC WHEN THEY EMERGE FROM THE RIBOSOMES (BY SIMILARITY).  
 CC -1- SUBUNIT: ARCHAEL SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA  
 CC MOLECULE OF 300 NUCLEOTIDES AND TWO PROTEIN SUBUNITS: SRP54 AND  
 CC SRP19 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- DOMAIN: HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE  
 CC M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE  
 CC SIGNAL SEQUENCE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AE005125; AAG20536.1;  
 DR InterPro: IPR008927; SRP54.  
 DR Pfam: PF00448; SRP54\_1.  
 DR PRODOM: PD000819; SRP54\_1.  
 DR PROSITE: PS00300; SRP54; FALSE\_NEG.  
 KW Signal recognition particle; GTP-binding; RNA-binding;  
 KW Complete proteome.  
 FT DOMAIN 1 289 G-DOMAIN (BY SIMILARITY).  
 FT DOMAIN 290 460 M-DOMAIN (BY SIMILARITY).  
 FT NP\_BIND 104 111 GTP (BY SIMILARITY).  
 FT NP\_BIND 184 188 GTP (BY SIMILARITY).  
 FT NP\_BIND 242 245 GTP (BY SIMILARITY).  
 FT DOMAIN 449 458 POLY-GLY.  
 SQ SEQUENCE 460 AA; 50149 MW; 8361B782651352E CRC64;

Query Match 22.3%; Score 52.5; DB 1; Length 460;  
 Best Local Similarity 42.1%; Pred. No. 30;  
 Matches 16; Conservative 3; Mismatches 18; Indels 1; Gaps 1;

OY 6 IDAIGSGADRAAEFDKMKKDDQIAAANVLRGMKDG 43  
 DB 215 LDAIGGAKQARQFQDASIGIDVAITK-LDGTAKGG 251

RESULT 8  
 ID SYA\_PYPAB STANDARD; PRT; 914 AA.  
 AC Q9UY36;  
 DT 20-AUG-2001 (rel. 40, Created)  
 DT 20-AUG-2001 (rel. 40, Last sequence update)  
 DE ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE-TRNA LIGASE) (ALARS).  
 GN ALAS OR PAB1245.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.  
 NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ORSAY;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
 RT structure and evolution.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +  
 CC PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ248288; CAB50576.1; -  
 DR InterPro: IPR002106; AA.tRNA\_Ligase\_II.  
 DR InterPro: IPR002318; tRNA-synt\_2c.  
 DR InterPro: IPR003156; DHHA1.  
 DR Pfam: PF02272; DHHA1; 1.  
 DR Pfam: PF02272; DHHA1; 1.  
 DR PRINTS: PR00980; TRNASYNTHALA.  
 DR PROSITE: PS00179; AA.tRNA\_LIGASE\_II\_1; FALSE\_NEG.  
 DR PROSITE: PS00339; AA.tRNA\_LIGASE\_II\_2; FALSE\_NEG.  
 DR AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 DR Complete proteome.  
 KW SEQUENCE 914 AA; 104858 MW; C4B449FE9E03B4E1 CRC64;  
 SQ  
 OY 6 IDAIGSGADRNAEAFDKMKDDOIAAAMVLRGMKDCQFAL 47  
 Db 825 IGEIVGSMDLRREAVRERKKPNRI---VVL--VSKEGFAL 861  
 Query Match 22.3%; Score 52.5; DB 1; Length 914;  
 Best Local Similarity 35.7%; Pred. No. 62;  
 Matches 15; Conservative 11; Mismatches 11; Indels 5; Gaps 2;  
 RESULT 9  
 PCNA\_SARCR STANDARD; PRT; 260 AA.  
 ID PCNA\_SARCR  
 AC O16852;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) (CYCLIN).  
 GN PCNA.  
 OS Sarcophaga crassipalpis.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Polyptera; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Oestroidea; Sarcophagidae; Sarcophaga.  
 OX NCBI\_TaxID=59312;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96382535; PubMed=9714841;  
 RA Tammaritello S.P., Denlinger D.L.;  
 RT "Cloning and sequencing of proliferating cell nuclear antigen (PCNA)  
 RT from the flesh fly, Sarcophaga crassipalpis, and its expression in  
 RT response to cold shock and heat shock.";  
 R1 Gene 215.425-429(1998)  
 CC -1- FUNCTION: THIS PROTEIN IS AN AUXILIARY PROTEIN OF DNA POLYMERASE  
 CC DELTA AND IS INVOLVED IN THE CONTROL OF EUKARYOTIC DNA REPLICATION  
 CC BY INCREASING THE POLYMERASE'S PROCESSIBILITY DURING ELONGATION  
 CC OF THE LEADING STRAND.  
 CC -1- SUBUNIT: HOMOTRIMER. FORMS A COMPLEX WITH ACTIVATOR 1  
 CC HEMEROPTAMER IN THE PRESENCE OF ATP (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE PCNA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF020427; AAC24238.1; -  
 DR HSP: P12004; IAXC.  
 DR InterPro: IPR000730; PCNA.  
 DR Pfam: PF00705; PCNA; 1.  
 DR PRINTS: PR00339; PCNACVCLIN.  
 DR ProDom: PD002673; PCNA; 1.

DR PROSITE: PS01251; PCNA\_1; 1.  
 DR PROSITE: PS00293; PCNA\_2; 1.  
 KW DNA-binding; Nuclear protein; DNA replication.  
 FT DNA\_BIND 61  
 FT SEQUENCE 260 AA; 28956 MW; 4645B748EB2996A8 CRC64;  
 SQ  
 OY 12 GSADRNAEAFDKMKDDOIAAAMVLRGMK 41  
 Db 48 GSTLRSDGFDFKRDNRISMGMLGSMK 77  
 Query Match 21.7%; Score 51; DB 1; Length 260;  
 Best Local Similarity 40.0%; Pred. No. 25;  
 Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;  
 RESULT 10  
 TRPD\_BRAJA STANDARD; PRT; 337 AA.  
 ID TRPD\_BRAJA  
 AC P94326;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18).  
 GN TRPD.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bradyrhizobium group; Bradyrhizobium.  
 OX NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=97214643; PubMed=9061023;  
 RA Kuykendall L.D., Hunter W.J.;  
 RT "The sequence of a symbolically essential Bradyrhizobium japonicum  
 RT operon consisting of trpd, trpc and a meac-like gene.";  
 RL Biochim. Biophys. Acta 1350:277-281(1997).  
 CC -1- CATALYTIC ACTIVITY: ANTHRANILATE + PHOSPHORIBOSYLTRYPOPHOSPHATE -  
 CC N-5'-PHOSPHORIBOSYL-ANTHRANILATE + PYROPHOSPHATE.  
 CC -1- PATHWAY: SECOND STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
 CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U79771; AAB39009.1; -  
 DR InterPro: IPR003262; Anthr.\_phosphoribosyltransf.  
 DR InterPro: IPR00312; Glycos.\_transf\_3.  
 DR InterPro: IPR000053; Thymid.phosphos.  
 DR Pfam: PF00591; Glycos.\_transf\_3\_1.  
 DR ProDom: PD001864; Glycos.\_transf\_3\_1.  
 DR ProDom: PD005916; Thymid.\_phosphos\_1.  
 KW Tryptophan biosynthesis; Transferase; Glycosyltransferase.  
 SQ SEQUENCE 337 AA; 34528 MW; C486203E9E5534E CRC64;  
 OY 11 GSGADRNAEAFDKMKDD-----QIAAMVLRGMK 42  
 Db 266 GGDADNMAIALOSVILNGKPSATRDVALMNAALVYAGRAKD 307  
 Query Match 21.7%; Score 51; DB 1; Length 337;  
 Best Local Similarity 35.7%; Pred. No. 33;  
 Matches 15; Conservative 3; Mismatches 14; Indels 10; Gaps 1;  
 RESULT 11  
 SUCC\_ECOLI

ID SUCC\_ECOLI STANDARD: PRT: 388 AA.  
AC P07460:  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SUCCINYL-COA SYNTHETASE BETA CHAIN (EC 6.2.1.5) (SCS-BETA).  
GN SUCC OR B0728 OR Z0882 OR ECS0753.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae:  
OC Escherichia  
OC NCBI\_TaxID=562, 83334;  
OX [1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86104124; PubMed=3002435;  
RA Buck D., Spencer M.E., Guest J.R.;  
RT "Primary structure of the succinyl-CoA synthetase of Escherichia  
RT coli.";  
RL Biochemistry 24:6245-6252(1985).  
RN [12]  
RP SEQUENCE FROM N.A.  
RX STRAIN-K12 / MG1655;  
MEDLINE=97426617; PubMed=9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [13]  
RP SEQUENCE FROM N.A.  
RX STRAIN-K12;  
MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajiura M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-153(1996).  
RN [14]  
RP SEQUENCE FROM N.A.  
RX STRAIN-O157:H7 / EDL933 / ATCC 700927;  
MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Plaut R.G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blatter F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
RN [15]  
RP SEQUENCE FROM N.A.  
RX STRAIN-O157:H7 / RIMD 0509952;  
MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsuka E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hattori M., Shingawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
RN [6]  
RP SEQUENCE OF 1-12.  
RX STRAIN-K12 / EMG2;  
MEDLINE=97443975; PubMed=9298646;  
RA Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
RT in the genome of Escherichia coli K-12.";  
RL Electrophoresis 18:1259-1313(1997).

RN [17]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=94193797; PubMed=8144675;  
RA Wolodko W.T., Fraser M.E., James M.N.G., Bridger W.A.;  
RT "The crystal structure of succinyl-CoA synthetase from Escherichia  
RT coli at 2.5-A resolution.";  
RL J. Biol. Chem. 269:10883-10890(1994).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=99141407; PubMed=9917402;  
RA Fraser M.E., James M.N., Bridger W.A., Wolodko W.T.;  
RT "A detailed structural description of Escherichia coli succinyl-CoA  
RT synthetase.";  
RL J. Mol. Biol. 285:1633-1653(1999).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
RX MEDLINE=20092606; PubMed=10625475;  
RA Joyce M.A., Fraser M.E., James M.N., Bridger W.A., Wolodko W.T.;  
RT "ADP-binding site of Escherichia coli succinyl-CoA synthetase  
RT revealed by X-ray crystallography.";  
RL Biochemistry 39:17-25(2000).  
CC -1- CATALYTIC ACTIVITY: SUCCINATE + COA + ATP -> SUCCINYL-COA + ADP +  
CC ORTHOPHOSPHATE.  
CC -1- ENZYME REGULATION: EXHIBITS TWO INTERESTING PROPERTIES: "SUBSTRATE  
CC SYNERGISM", IN WHICH THE ENZYME IS MOST ACTIVE FOR THE CATALYSIS  
CC OF ITS PARTIAL REACTIONS ONLY WHEN ALL THE SUBSTRATE BINDING SITES  
CC ARE OCCUPIED, AND "CATALYTIC COOPERATIVITY" BETWEEN ALTERNATING  
CC ACTIVE SITES IN THE TETRAMER, WHEREBY THE INTERACTION OF  
CC SUBSTRATES (PARTICULARLY ATP) AT ONE SITE IS NEEDED TO PROMOTE  
CC CATALYSIS AT THE OTHER.  
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.  
CC -1- MISCELLANEOUS: SUCCINYL-COA SYNTHETASE (SCS) OF E. COLI CATALYZES  
CC ITS REACTION VIA THREE STEPS THAT INVOLVE PHOSPHORYL ENZYME AND  
CC ENZYME-BOUND SUCCINYL PHOSPHATE AS INTERMEDIATES.  
CC -1- MISCELLANEOUS: DURING AEROBIC METABOLISM IT FUNCTIONS IN THE  
CC CITRIC ACID CYCLE, COUPLING THE HYDROLYSIS OF SUCCINYL-COA TO THE  
CC SYNTHESIS OF ATP & THUS REPRESENTS AN IMPORTANT SITE OF  
CC SUBSTRATE-LEVEL PHOSPHORYLATION. IT CAN ALSO FUNCTION IN THE  
CC OTHER DIRECTION FOR ANABOLIC PURPOSES, AND THIS MAY BE  
CC PARTICULARLY IMPORTANT FOR PROVIDING SUCCINYL-COA DURING ANAEROBIC  
CC GROWTH WHEN THE OXIDATIVE ROUTE FROM 2-OXOGLUTARATE IS SEVERELY  
CC REPRESSED.  
CC -1- MISCELLANEOUS: THE BETA-SUBUNIT CONTAINS THE ATTACHMENT SITES FOR  
CC SUCCINATE. THE COMPLETE ACTIVE SITE IS PROBABLY LOCATED IN THE  
CC REGION OF ALPHA- BETA CONTACT.  
CC -1- SIMILARITY: TO OTHER BETA SUBUNITS OF SUCCINYL-COA SYNTHETASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: J01619; AAA23899.1; -;  
CC EMBL: AE000176; AAC38822.1; -;  
CC EMBL: D90711; BA35394.1; -;  
CC EMBL: AE005251; AAG55052.1; -;  
CC EMBL: AP002553; BAB34176.1; -;  
CC PIR: A24090; SYEC5B.  
CC PDB: 1SCU; 20-APR-95.  
CC PDB: 28CU; 02-AUG-99.  
CC PDB: 1COT; 10-JAN-00.  
CC PDB: 1COJ; 10-JAN-00.  
CC SWISS-2DPAGE: P07460; COLI.  
CC ECO2DBASE: E039.8; 6TH EDITION.  
CC EcoGene: EG10981; succ.  
CC InterPro: IPR003135; ATP-grasp.  
CC InterPro: IPR000303; CoA\_ligase.  
CC Pfam: PF02222; ATP-grasp; 1.  
CC Pfam: PF00549; ligase-CoA; 1.  
CC PROSITE: PS01217; SUCCINYL\_COA\_LIG\_3; 1.

KW Ligase: Tricarboxylic acid cycle: 3D-structure: complete proteome.  
SQ SEQUENCE 388 AA: 41392 MW: 09C429EC97A823CF CRC64:

Query Match 21.7%; Score 51; DB 1; Length 388;  
Best Local Similarity 34.4%; Pred. No. 39;  
Matches 11; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

OY 4 NPIDALIGSADRNAEAFDKMKKDDQIAAAMY 35  
DB 285 NPLDVGATKERTVETAFKILISDKAVAVLV 316

## RESULT 12

P39\_BRUAB STANDARD; PRT; 401 AA.  
AC 006875;

DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
E IMMUNOGENIC 39 KDA PROTEIN.

CC Brucella abortus.

OC Bacteria: Proteobacteria; alpha subdivision: Rhizobiaceae group;

OC Brucellaceae: Brucella.

OX NCBI\_TaxID=235;

RN [1]  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC STRAIN=544;  
RX MEDLINE=97162316; PubMed=9009303;

RA Demol P.A., Vo T.K., Tibor A., Weynants V.E., Trunde J.M., Dubray G.,  
RT Linet J.N., Letesson J.J.;  
RT "Characterization, occurrence, and molecular cloning of a  
RT 39-kilodalton Brucella abortus cytoplasmic protein immunodominant in  
RT cattle.";

RL Infect. Immun. 65:495-502(1997).

CC -1- FUNCTION: ALLEGENIC IN CATTLE.

CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
CC PROTEIN FAMILY 1.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L35038; AAB50572.1; -;  
CC InterPro: IPR000567; SBP\_bac\_1.

DR Pfam: PF01547; SBP\_bacterial\_1; 1.  
DR PROSITE: PS01037; SBP\_BACTERIAL\_1; FALSE-NEG.

KW Allergen: Transport.

SQ SEQUENCE 401 AA: 42901 MW: 701470A378F338FE CRC64:

Query Match 21.7%; Score 51; DB 1; Length 401;  
Best Local Similarity 34.1%; Pred. No. 40;  
Matches 15; Conservative 7; Mismatches 8; Indels 14; Gaps 2;

OY 9 AIGGSADRNAEAFDKMKK-----DDQIAAAMYLKRGMA 40  
DB 203 ALG--SDTWKQAFDRMSKRLRYVDNFGSRDWNLASAMVIEGKA 244

## RESULT 13

SYM\_ORYSA STANDARD; PRT; 804 AA.  
AC 092751;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA

DE LIGASE) (METRS).

OS Oryza sativa (Rice).

OC Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]  
RP SEQUENCE FROM N.A.

RA Denizliak M., Miranda M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +  
CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF040700; AAC99620.1; -;  
CC InterPro: IPR002300; trna-synt\_1a.

DR InterPro: IPR001412; trna-synt\_1a.

DR InterPro: IPR002304; trna-synt\_met.

DR InterPro: IPR002547; trna-bind.

DR Pfam: PF00133; trna-synt\_1; 1.

DR Pfam: PF01588; trna-bind\_1.

DR PRINTS: PR01041; TRNASYNTHET.

DR PROSITE: PS00178; AA-TRNA\_LIGASE\_1; 1.

KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW trna-binding.

FT DOMAIN 4 11 POLY-PRO.  
FT SITE 28 38 "HIGH" REGION.  
FT SITE 350 354 "KMSKS" REGION.  
FT BINDING 353 353 ATP (BY SIMILARITY).  
FT DOMAIN 648 743 TRNA-BINDING.  
SQ SEQUENCE 804 AA: 89716 MW: A87E8ABBD419D440 CRC64:

Query Match 21.7%; Score 51; DB 1; Length 804;  
Best Local Similarity 35.0%; Pred. No. 84;  
Matches 14; Conservative 9; Mismatches 13; Indels 4; Gaps 1;

OY 2 LTNPIDALIGSADRNA---EAFDKMKRDDQIAAAMVLR 37  
DB 465 LKQGLKSMGICSSDGNAYLQESQFMKLYKEDPRACAVVWK 504

## RESULT 14

PSMA\_METTE STANDARD; PRT; 247 AA.  
AC 059565;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROTEASOME ALPHA SUBUNIT (EC 3.4.99.46) (MULTICATALYTIC ENDOPEPTIDASE  
DE COMPLEX ALPHA SUBUNIT).

GN PSMA.

OS Methanosarcina thermophila.  
OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;  
OC Methanosarcina.

OX NCBI\_TaxID=2210;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1825 / TM-1;  
RX MEDLINE=96081920; PubMed=7499378;  
RT Maupin-Furlow J.A., Ferry J.G.;  
RT "A proteasome from the methanogenic archaeon Methanosarcina  
RT thermophila";  
RT J. Biol. Chem. 270:28617-28622(1995).

-1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR SLIGHTLY BASIC PH (BY SIMILARITY).  
 CC -1- PATHWAY: INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL PROTEOLYTIC PATHWAY.  
 CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS, ALPHA AND BETA. THE COMPLEX IS FORMED OF FOUR RINGS. THE TWO OUTER RINGS ARE EACH COMPOSED OF SEVEN ALPHA SUBUNITS. THE TWO INNER RINGS ARE EACH COMPOSED OF SEVEN BETA SUBUNITS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE PROTEASOME A-TYPE FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U30483; AAA93166.1; -  
 DR InterPro: IPR001353; Proteasome.  
 DR InterPro: IPR000426; Proteasome\_A.  
 DR Pfam: PF00227; Proteasome; 1.  
 DR PROSITE: PS00388; PROTEASOME\_A; 1.  
 DR Proteasome; Hydrolase; Protease.  
 SQ SEQUENCE 247 AA; 27155 MW; B6897277A3179A5A CRC64;

Query Match 21.3%; Score 50; DB 1; Length 247;  
 Best Local Similarity 30.9%; Pred. No. 32;  
 Matches 17; Conservative 7; Mismatches 19; Indels 12; Gaps 2;  
 QY 3 TNPIDA-----AIGGSADRNAEAFDKMKDDQIAAMVLRGM-----AKDGQF 45  
 Db 150 TDPGALLLEYKATATAGRNAYVEFADYKEDMNEAAILGMDALYKRAEGKF 204

RESULT 15  
 Y024\_MYCPN STANDARD; PRT; 362 AA.  
 AC P75088;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 FT PROBABLE GTP-BINDING PROTEIN MG024 HOMOLOG (B01\_ORF362).  
 G MPN026 OR MP128.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;  
 OC Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelfeich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Herrmann R.;  
 RT \*Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.\*  
 RT Nucleic Acids Res. 24:4420-4449(1996).  
 RL -1- SIMILARITY: BELONGS TO THE YCHF FAMILY OF GTP-BINDING PROTEINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000015; AAB95776.1; -

DR InterPro: IPR000765; GTP1\_OBG.  
 DR Pfam: PF01018; GTP1\_OBG; 1.  
 DR PRINTS: PR00326; GTP1OBG.  
 KW GTP-binding; Complete proteome.  
 FT NP\_BIND 8 15 GTP (POTENTIAL).  
 FT NP\_BIND 71 75 GTP (POTENTIAL).  
 SQ SEQUENCE 362 AA; 40607 MW; 7C79B46C84D5FEF83 CRC64;

Query Match 21.3%; Score 50; DB 1; Length 362;  
 Best Local Similarity 35.5%; Pred. No. 48;  
 Matches 11; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 12 GSADRNAEAFDKMKRRDDQIAAMVLRGMKD 42  
 Db 144 GKIKKRAESGDQSKKEFYGLAPVIGLQGN 174

Search completed: January 10, 2002, 14:09:58  
 Job time: 327 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 40.15 Seconds  
(without alignments)  
643.875 Million cell updates/sec

Title: US-09-445-803-2

Perfect score: 1708  
Sequence: 1 KNNDDHNMHGTVKNAVDMAK.....GNGATKAGDASVNCIANG 349

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

T number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_1101:\*

1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT:\*

2: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT:\*

3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT:\*

4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT:\*

5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT:\*

6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT:\*

7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT:\*

8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT:\*

9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT:\*

10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT:\*

11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT:\*

12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT:\*

13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT:\*

14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT:\*

15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT:\*

16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:\*

17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT:\*

18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:\*

19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:\*

20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:\*

21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1708	100.0	349	20	AAW95612
2	1695	99.2	349	21	AAB36281
3	596	34.9	1328	20	AAV20088
4	538.5	31.5	533	20	AAV20112
5	437.5	25.6	356	18	AAW2676
6	345.5	20.2	156	20	AAV20113
7	222	13.0	168	20	AAV20089
8	220.5	12.9	738	19	AAW56163
9	214	12.5	681	22	AAAB2609
10	210.5	12.3	291	22	AAAB2608
11	208.5	12.2	528	22	AAAB2611

12	207	12.1	691	22	AAAB2610	Spider recombinant
13	203.5	11.9	604	16	AAAB9057	Spider dragline va
14	203.5	11.9	606	16	AAAB9055	Spider dragline va
15	203.5	11.9	606	20	AAV40101	Polymer of an anal
16	203.5	11.9	606	20	AAV40102	Polymer of an anal
17	197	11.5	651	20	AAV40097	Spider silk protei
18	197	11.5	718	12	AAV4308	N.clavipes draglin
19	197	11.5	718	19	AAV5346	Nephila clavipes s
20	197	11.5	718	21	AAV59070	N.clavipes spider
21	196.5	11.5	831	16	AAAB0168	PMISSI MSP spider
22	194.5	11.4	606	16	AAAB9053	Spider dragline va
23	194.5	11.4	606	20	AAV40100	Polymer of an anal
24	194	11.4	615	20	AAV40099	Spider silk protei
25	193.5	11.3	429	10	AAAB0064	Antigenic protein
26	193	11.3	344	17	AAAB08097	Vmp7 soluble varia
27	189.5	11.1	646	18	AAW27178	Nephila clavipes s
28	187	10.9	898	18	AAW31853	Mycobacterium tube
29	184.5	10.8	1488	22	AAAG1741	C glutamicum prote
30	178	10.4	456	21	AAV67238	Amino acid sequenc
31	177.5	10.4	1822	13	AAAB27745	Extracellular fact
32	177	10.4	2411	21	AAAB2860	Haemophilus influe
33	176.5	10.3	47	20	AAW95613	Borrelia burgdorfe
34	176.5	10.3	402	8	AAV70709	Plasmodium cynomol
35	172.5	10.1	261	19	AAW79137	FlGA Gly-ala inser
36	171.5	10.0	641	20	AAV28843	Epostein Barr Virus
37	171.5	10.0	641	21	AAV5856	Epostein Barr Virus
38	171.5	10.0	641	22	AAAB2332	EBV tethering prot
39	170.5	10.0	1018	16	AAW01496	Silk-like protein
40	170.5	10.0	1038	17	AAAB95107	Fibronectin cell b
41	170.5	10.0	1038	22	AAAB72727	Repetitive protein
42	170.5	10.0	1038	22	AAAB63997	PCB-SLP protein fr
43	169.5	9.9	3596	21	AAV87407	Bordetella pertuss
44	169.5	9.9	3647	11	AAV05041	Filamentous haemag
45	169	9.9	1095	21	AAAB01835	Haemophilus influe

#### ALIGNMENTS

RESULT 1

AAW95612

ID AAW95612 standard; Protein: 349 AA.

AC AAW95612;

XX

XX

08-JUN-1999 (first entry)

DT

XX

DE Borrelia burgdorferi surface antigen P39.5 clone 7-1 polypeptide.

XX

XX

KW Lyme disease: surface antigen; P39.5; diagnosis; prevention;

KW vaccine; antisense; therapy; treatment; primer; probe;

KW antibody; DNA.

XX

XX

OS Borrelia burgdorferi.

XX

XX

PM W09900413-A1.

XX

XX

07-JAN-1999.

PD

XX

XX

29-JUN-1998; 98WO-US13551.

PE

XX

XX

30-JUN-1997; 97US-0051271.

PR

XX

XX

(TULANE ) TULANE EDUCATIONAL FUND.

PA

XX

PI Philipp MT;

XX

XX

WPI: 1999-095676/08.

DR

XX

XX

N-PSDB: AAX07411.

PT

PT New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi

PT - and related vectors, transformants, antibodies and polypeptides,

PT for diagnosis, prevention and treatment of Lyme disease

xx Claim 5; Pages 54-55; 89pp; English.  
ps The sequence is that of a Borrelia burgdorferi surface antigen P39.5  
cc Clone 7-1 polypeptide. It can be used to raise antibodies to, and  
cc or fragments of it which may be used in the production of P39.5  
cc In the development of vaccines against Lyme disease. The sequence  
cc can also be used for making primers and probes for diagnosis, also  
cc In DNA vaccines, as antisense therapeutics and for drug screening.  
cc Antibodies can be used as diagnostic (immunoassay) reagents, for  
cc treating Lyme disease, for affinity purification, for drug  
cc screening and to produce anti-idiotypic antibodies (used in the  
cc same way as P39.5 to induce an immune response).  
xx Sequence 349 AA;  
SQ

Query Match 100.0%; Score 1708; DB 20; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1.2e-110;  
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KNDHDNHKGTVKNAVDMAKAAEEAASAAATGNAATGIVYKNSGAAGGEAASVNGI 60  
DB 1 knndhdnhkgtvknaavdmakaaeeasaatgnaaigdvvknsgeaakggeaasvngi 60  
OY 61 AKGIGIVDAAGKADAKEGKLDATGAEETTNNAGKLFYKRAADGADADGAKAAAVA 120  
DB 61 akgigivdaagkadegekldatgaegttnnagklfvykraadggadddgkkaava 120  
OY 121 ASAAATGNAATGIVYNGDVAKAKAGDAASVNGIAKIGIVDAAEKADAKEGKLNAAAGAG 180  
DB 121 asaatgnaaigdvngdvakakagdaasvngiakgigivdaaekadegeklnaaagag 180  
OY 181 TTNADAGKLFVKNAGNNGGEGADGAKAAAVAAVSGEOLIKAIIVHAAKDGEGKCKRAAD 240  
DB 181 ttnadagklfvknagnnggeadgakaavaavsgelikaivhaakdgegkckrkaad 240  
OY 241 RPNPIDAATGAGDNDAAAFATMKKDDQIAAAMVLRGMKDGOPALDAAAHGCTYKN 300  
DB 241 rpnpidaatgagdnadaaafatmkkddqiaaamvlgmkgdgfalxdaaaahgctvkn 300  
OY 301 AVDIITKAAAEASAAATGSAATGIVYNGATGATKAGDAKSVNGIANG 349  
DB 301 avdiitkaaeasaaatgsaatgsvaigdvngngatagdgdaeksvngiakg 349

## RESULT 2

AAAB36281 standard; peptide; 349 AA.  
AAB36281;

21-FEB-2001 (first entry)

B. garinii P7-1 protein.

Variable surface antigen; invariable region; VISE; Lyme disease;  
Lyme borreliosis.

Borrelia garinii.

WO200065064-A1.

02-NOV-2000.

25-APR-2000; 2000WO-US11085.

28-APR-1999; 99US-0300971.

(TULANE ) TULANE EDUCATIONAL FUND.

Philipp MT, Liang FT;

DR WPI: 2000-687350/67.  
xx Novel peptides comprising an invariable 26-amino acid long region  
PT isolated from Borrelia burgdorferi (sensu lato), useful for rapid and  
PT specific diagnosis of Lyme disease.  
xx Example 2; Fig 2; 76pp; English.  
xx The present invention describes several peptides comprised of the  
CC invariable regions IRL-IR6 of the B. burgdorferi variable surface antigen  
CC (VISE) variable domain. These peptides can be used in the specific  
CC diagnosis, treatment and vaccination against B. burgdorferi, which causes  
CC Lyme disease (also known as Lyme borreliosis) in humans, dogs, horses,  
CC cows and other animals.  
xx Sequence 349 AA;  
SQ

Query Match 99.2%; Score 1695; DB 21; Length 349;  
Best Local Similarity 99.4%; Pred. No. 9.3e-110;  
Matches 347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KNDHDNHKGTVKNAVDMAKAAEEAASAAATGNAATGIVYKNSGAAGGEAASVNGI 60  
DB 1 knndhdnhkgtvknaavdmakaaeeasaatgnaaigdvvknsgeaakggeaasvngi 60  
OY 61 AKGIGIVDAAGKADAKEGKLDATGAEETTNNAGKLFYKRAADGADADGAKAAAVA 120  
DB 61 akgigivdaagkadegekldatgaegttnnagklfvykraadggadddgkkaava 120  
OY 121 ASAAATGNAATGIVYNGDVAKAKAGDAASVNGIAKIGIVDAAEKADAKEGKLNAAAGAG 180  
DB 121 asaatgnaaigdvngdvakakagdaasvngiakgigivdaaekadegeklnaaagag 180  
OY 181 TTNADAGKLFVKNAGNNGGEGADGAKAAAVAAVSGEOLIKAIIVHAAKDGEGKCKRAAD 240  
DB 181 ttnadagklfvknagnnggeadgakaavaavsgelikaivhaakdgegkckrkaad 240  
OY 241 RPNPIDAATGAGDNDAAAFATMKKDDQIAAAMVLRGMKDGOPALDAAAHGCTYKN 300  
DB 241 rpnpidaatgagdnadaaafatmkkddqiaaamvlgmkgdgfalxdaaaahgctvkn 300  
OY 301 AVDIITKAAAEASAAATGSAATGIVYNGATGATKAGDAKSVNGIANG 349  
DB 301 avdiitkaaeasaaatgsaatgsvaigdvngngatagdgdaeksvngiakg 349

## RESULT 3

AAAY20088 standard; Protein; 1328 AA.  
AAAY20088;

AAV20088;

19-JUL-1999 (first entry)

B. burgdorferi antigenic protein, f24-1.aa.

Antigenic protein; vaccine; Lyme disease; infection; detection.

Borrelia burgdorferi.

WO9859071-A1.

30-DEC-1998.

18-JUN-1998; 98WO-US12718.

03-SEP-1997; 97US-0057483.

20-JUN-1997; 97US-0050359.

22-JUL-1997; 97US-0053344.

22-JUL-1997; 97US-0053377.  
(HUMA-) HUMAN GENOME SCI INC.



AAW2676 standard; Protein; 356 AA.

AAW2676:

22-FEB-1998 (first entry)

Borrelia variable major protein (VMP)-like protein VlsE.

Variable major protein-like sequence: VMP-like sequence; vls locus; Lyme disease; relapsing fever; therapy; diagnosis; vaccine.

Borrelia burgdorferi strain B31-5A3 (ATCC 35210).

Key Location/Qualifiers

Peptide 1..19

/label= Sig-peptide

/note= "lipoprotein signal peptide"

MOW9731123-AI.

28-AUG-1997.

20-FEB-1997: 97MO-USO2952.

21-FEB-1996: 96US-0012028.

(TEXA ) UNIV TEXAS SYSTEM.

Barbour AG, Hardham JM, Howell JK, Norris SJ, Weinstein GM; Zhang J;

WPJ: 1997-435172/40.

N-PsDB: AAT85042.

Nucleic acid encoding variable major protein-like peptide of Borrelia - useful for recombinant production of VMP like protein or peptide, or for diagnosis of Lyme disease

Claim 2: Page 97-99; 130pp; English.

This protein comprises the surface-exposed lipoprotein variable major protein (VMP)-like protein VlsE of Borrelia burgdorferi. Its sequence was deduced from an isolated VlsE gene (see AAT85042). An elaborate genetic system has been identified that promotes extensive antigenic variation of VlsE. An infectivity related 28-kb linear plasmid, pBB28la, of B. burgdorferi B31 contains a vmp-like sequence (vls) locus consisting of 15 silent vls cassettes (see AAT85043) and the expressed VlsE gene. Portions of several of the 15 silent vls cassette sequences, located approx. 500 bp upstream of VlsE, recombine into the central VlsE cassette region during infection, resulting in antigenic variation and hence immune evasion, long-term survival and pathogenesis in the mammalian host. Recombinant or native proteins expressed by VMP-like genes, will be useful for the immunotherapy, immunophylaxis and immunodiagnosis of Lyme disease, relapsing fever and related disorders in humans and animals. They can also be used in a Lyme disease vaccine.

Sequence 356 AA:

Query Match	25.68;	Score 437.5;	DB 18;	Length 356;
Best Local Similarity	46.98;	Pred. No. 5.6e-23;		
Matches 127;	Conservative 26;	Mismatches 87;	Indels 31;	Gaps 11.

QY 79 GKLDATGAEGTTNVNAGKLEVKRAADDGGDADDAGKAAAAVAASAATGNAIGDVVNGDV 138

Db 103 gkpdstgsvgt-----vegaIkeVseIIdklVkvktaegassgtaigev-ada 153

QY 139 AKAKGDAASVNGIAKGIKIVDAEAKADAKEGKLN - AGAEGTTNADAGKLFVKNAGNV 197

Db 154 daakvadkasvkgiakgikeiveaagse---klkavaaakgenkagaklfkgagaaa 209

1

```

Oy 198 GGAGNDACKAAAATAAASVSEBILATVYHAADGGGKOKKAAADNTNIDMAIGSAGND 257
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 hgdseaeaaakgaagsavsgeqllsaiytaa--daaeqdgkpkpeekknplaaia--gqkdg 265
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 258 AAAFA--TMRKDDQDIAAMVLRGMAKKGQFALKTDAAAHAECTVNAVDIIRKAAEEA--S 313
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 266 gaefgqdmkxkddqilaaiialtrymakdgkfvkv-----gkakeagallgaeeavrk 319
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 314 AASAAATGSAATGDVYNGNGATFAKGDGAKSVN 344
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 320 vlgaltg--lllgdavs--sgllkvdyvxaaas 347
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT	6
AAV20113	
ID	AAV20113 standard; Protein; 156 AA

DT 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein, t49-2.aa.

**KW** Antigenic protein; vaccine; Lyme disease; infection; detection.

OS Borrelia burgdorferi.

PN WO9859071-A1.

PD 30-DEC-1998.

AA 18-JUN-1998; 98WO-US12718  
PF

AA  
PR 03-SEP-1997; 97US-0057483.

PR	20-JUN-1997;	97US-0050359
PR	22-JUL-1997;	97US-0053344

PR 22-JUL-1997; 9705-0053377.  
XX XX

PA (HUMA-) HUMAN GENOME SCI INC  
PA (MEDI-) MEDIMTNE INC

XX  
XX  
Cboj CH  
Evdja AT  
Hacccc MC  
Lethjexa B

XX  
1000-190090/15

DR N-PSDB; AAX61810.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop

PT caused by Borrelia, particularly Lyme disease

PS Claim 12; Page 202; 275pp; English

aa This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC

invention, which is suitable for use in a vaccine. The BB polypeptides can be used in vaccines for eliciting protective antibodies to members of

CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an

CC Infection caused by a member of the *Borrelia* genus. The products can also  
CT be used for detection of members of the *Borrelia* genus

XX  
XX  
Sequence 156 AA.

Query Match	Score	DB	Length
20.28;	345.5;	20;	156;

Matches	86;	Conservative	17;	Mismatches	51;	Indels	7;	Gaps	4;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

QY 140 KAKGDA SVNGIAKIGI VDAE KADAKEGLNAAGAE GTNADAGKLFVKNAGNVGG 199

Db 2 ksqvadkasvtgiakgikeli veaagse---kikvaagegenekagklfgkagagnag 57

QY 200 EAGDAGKAAAVAAVSGEQLKATVHAAKDGEKQGGKKAADRTNPIDAAIGGAGDNDAA 259

[illegible]





RESULT 9  
 AAB82609 standard; Protein; 681 AA.  
 ID AAB82609 standard; Protein; 681 AA.  
 AC AAB82609;  
 XX  
 DT 02-OCT-2001 (first entry)  
 XX  
 DE Spider recombinant silk protein pOE((SP1)4/(SP2)1)4.  
 XX  
 KW Spider: orb-weaver; silk protein; pOE((SP1)4/(SP2)1)4;  
 XX structural protein; purification; fibre; spinning.  
 OS Nephila clavipes.  
 XX  
 PN MO20015333-A1.  
 XX  
 XX 26-JUL-2001.  
 PD  
 XX 01-NOV-2000; 2000MO-US30086.  
 XX  
 PR 20-JAN-2000; 2000US-0490291.  
 XX  
 PA (MELL/) MELLO C M.  
 PA (ARCI/) ARCIDIACONO S.  
 PA (BUTL/) BUTLER M M.  
 PA (USSA ) US SEC OF ARMY.  
 XX  
 PI Mello CM, Arcidiacono S, Butler MM;  
 XX  
 DR WP1: 2001-483136/52.  
 DR N-PSDB; AAH26302.  
 PT  
 PT Recovering structural polypeptides in a biological sample, useful for  
 purifying and spinning spider silks and other structural proteins,  
 PT comprises treating the sample containing the polypeptides with an acid  
 solution.

Claim 2; Page 34-37; 49pp; English.

The present sequence is that of orb-weaver spider (Nephila clavipes)  
 recombinant silk protein pOE((SP1)4/(SP2)1)4. The invention  
 provides methods for purifying and spinning spider silks and other  
 structural proteins. Organic acids are used to lyse recombinant  
 cells or other biological samples (such as non-recombinant  
 derived cells), and enrich the purity and yields of structural  
 proteins by hydrolysing many of the macromolecules while leaving  
 the structural proteins intact. In the case of silk proteins, the  
 resulting lysate is further purified by ion-exchange or affinity  
 chromatography and processed into an aqueous-based mixture for  
 fibre spinning. In the present case, the pOE((SP1)4/(SP2)1)4 gene  
 was cloned into vector pOE-9 for recombinant expression in  
 Escherichia coli, and recombinant silk protein was obtained in 9%  
 purity using propionic acid and anion-exchange chromatography on  
 OAE-Sephadex A50. Products obtained using the methods of the  
 invention can be used in the construction of many materials  
 including films, fibres, woven articles, sutures, ballistic  
 protection, parachutes and parachute cords. The new method has  
 the following advantages over prior art: it involves fewer steps,  
 requires less time and smaller volumes of reagents, results in  
 better recovery of protein at higher purity (70-99%), is easy to  
 scale up, and the fibres are spun in an environmentally benign  
 solution reducing hazardous waste accumulation and cost.

Sequence 681 AA:

Query Match	12.5%	Score 214;	DB 22;	Length 681;
Best Local Similarity	26.6%;	Pred. No. 3e-07;		
Matches	93;	Conservative 41;	Mismatches 151;	Indels 64;
			Gaps 15;	
OY	1 KNNDDHNKICVYKNVD--MAKADEEASASASATGNAIGCV--VKNSGAACKGCEAAS	56		
	: : : : : :	:	: : : : : :	: : : : : :

Db	2	rgshhhhhgsmasgryglgqggaagaaaaaaggaqggygglgsgqltsgrg-----	57
Qy	57	VNGIAKGIKGIIVDAAGRAADAKEKGLDATTGEGTTNVNAGKLFVKRAADGDGADDAAGKAA	116
Db	58	lgqgqgaagaaaaaaggaqggygglgsgqltsq-----	103
Qy	117	AAVAASAATGAAATIGDV--VNGDVAKAKGGDAASVNGIAKGIKGIIVDAEAKDAKEKGLN	174
Db	104	aaaaaaggaqggygglgsgqltsgrg---lgqggaagaaaaaaggaqggyg	159
Qy	175	AAGAGGTNNADA---GKLFVKNNAGNNGGEGAGDAGKAAAVAAVSGEQILKALVHAAKKG	231
Db	160	glsgqtsqggygpygsgq-tsrgyglgqgq--qgaagaaaaag-----	207
Qy	232	EKOGKKAADRTNPIDAIIGAGAGNDAAAFATKKKDDQIAAMAVLKGAKDGQFALKDAA	291
Db	208	-ygglgsgql-----sgrglgqggaagaaaa-----	253
Qy	232	AAHGTYNKNAVDTIIKAAAEASASAAATGSAATIGDV--VNGMGATPAKGG	338
Db	234	tsrgyglgq-----qgaagaaaaaaggaqggygglgsgqltsgrg	297
RESULT 10			
AA882608			
ID	AA882608	standard; Protein; 291 AA.	
XX			
AC	AA882608;		
XX			
DT	02-OCT-2001	(first entry)	
XX			
DE	Spider recombinant silk protein pQE(sp1)7.		
XX			
KW	Spider: orb-weaver; silk protein; pQE(sp1)7; structural protein;		
XX			
KW	purification; fibre; spinning.		
XX			
OS	Nephila clavipes.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference 232	/note- "encoded by GGT"	
XX			
PN	WO20015333-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	01-NOV-2000; 2000MO-US30086.		
XX			
PR	20-JAN-2000; 2000US-0490291.		
XX			
PA	(MELLO/) MELLO C M.		
XX	(ARCI/) ARCIDIACONO S.		
PA	(BUTL/) BUTLER M M.		
XX	(USSA ) US SEC OF ARMY.		
PI	Mello CM, Arcidiacono S, Butler MM;		
XX			
DR	WPI: 2001-483136/52.		
XX	N-PSDB: AAH26301.		
XX			
PT	Recovering structural polypeptides in a biological sample, useful for		
PT	purifying and spinning spider silks and other structural proteins,		
PT	comprises treating the sample containing the polypeptides with an acid		
XX			
PS	Claim 2; Page 33-34; 49pp; English.		
XX			
CC	The present sequence is that of orb-weaver spider (Nephila clavipes)		
CC	recombinant silk protein, pQE(sp1)7. The invention provides		
CC	methods for purifying and spinning spider silks and other		
CC	structural proteins. Organic acids are used to lyse recombinant		
CC	cells or other biological samples (such as non-recombinant		
CC	derived cells), and enrich the purity and yields of structural		









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 19.72 Seconds  
(without alignments)  
398.258 Million cell updates/sec

Title: US-09-445-803-2

Perfect score: 1708  
Sequence: 1 KNNDDHDKSTVKNAYDMK.....GNATKAGDAKSYNGIAGK 349

Scoring table: BIOSQM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	220.5	12.9	738	3	US-08-864-038A-3 Sequence 3, Appl1
2	207.5	12.1	606	4	US-09-247-806-6 Sequence 6, Appl1
3	203.5	11.9	604	4	US-08-556-978B-63 Sequence 63, Appl1
4	203.5	11.9	606	4	US-08-556-978B-23 Sequence 23, Appl1
5	203.5	11.9	606	4	US-09-247-806-8 Sequence 8, Appl1
6	197	11.5	651	4	US-08-556-978B-19 Sequence 19, Appl1
7	197	11.5	651	4	US-09-247-806-1 Sequence 1, Appl1
8	197	11.5	718	1	US-08-425-069-2 Sequence 2, Appl1
9	197	11.5	718	2	US-08-317-844B-2 Sequence 2, Appl1
10	197	11.5	747	3	US-09-034-177-3 Sequence 3, Appl1
11	196.5	11.5	832	1	US-08-209-747-2 Sequence 2, Appl1
12	196.5	11.5	832	1	US-08-458-298-2 Sequence 2, Appl1
13	194.5	11.4	606	4	US-08-556-978B-21 Sequence 21, Appl1
14	194.5	11.4	606	4	US-09-247-806-4 Sequence 4, Appl1
15	193	11.3	344	1	US-07-941-523-24 Sequence 24, Appl1
16	170.5	10.0	1038	1	US-07-609-716-36 Sequence 36, Appl1
17	170.5	10.0	1038	4	US-08-475-411A-36 Sequence 36, Appl1
18	170.5	10.0	1038	4	US-08-478-029A-36 Sequence 36, Appl1
19	163	9.5	235	2	US-08-529-190B-1 Sequence 1, Appl1
20	163	9.5	1018	1	US-08-089-862-11 Sequence 11, Appl1
21	163	9.5	1018	1	US-08-587-333-16 Sequence 16, Appl1
22	163	9.5	1018	5	PCT-US94-0776-16 Sequence 16, Appl1
23	163	9.5	1338	2	US-08-728-470-9 Sequence 9, Appl1
24	163	9.5	1338	4	US-08-719-641-9 Sequence 9, Appl1
25	163	9.5	1529	2	US-08-728-470-10 Sequence 10, Appl1
26	163	9.5	1529	4	US-08-719-641-10 Sequence 10, Appl1
27	163	9.5	1599	2	US-08-617-697-9 Sequence 9, Appl1

28	163	9.5	1600	2	US-08-617-697-10 Sequence 10, Appl1
29	162	9.5	766	1	US-08-175-155-53 Sequence 53, Appl1
30	162	9.5	766	1	US-08-477-509B-88 Sequence 88, Appl1
31	162	9.5	766	2	US-08-707-237A-61 Sequence 61, Appl1
32	162	9.5	766	3	US-08-482-085B-88 Sequence 88, Appl1
33	162	9.5	979	1	US-08-477-509B-89 Sequence 89, Appl1
34	162	9.5	979	3	US-08-482-085B-89 Sequence 89, Appl1
35	162	9.5	1050	1	US-08-175-155-54 Sequence 54, Appl1
36	161.5	9.5	595	1	US-08-425-069-4 Sequence 4, Appl1
37	161.5	9.5	595	2	US-08-317-844B-4 Sequence 10, Appl1
38	161	9.4	2123	4	US-08-968-685A-10 Sequence 31, Appl1
39	160	9.4	1177	1	US-07-609-716-31 Sequence 29, Appl1
40	160	9.4	1177	1	US-08-175-155-29 Sequence 64, Appl1
41	160	9.4	1177	1	US-08-477-509B-64 Sequence 35, Appl1
42	160	9.4	1177	2	US-08-707-237A-35 Sequence 64, Appl1
43	160	9.4	1177	3	US-08-482-085B-64 Sequence 31, Appl1
44	160	9.4	1177	4	US-08-475-411A-31 Sequence 31, Appl1
45	160	9.4	1177	4	US-08-478-029A-31 Sequence 31, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-864-038A-3  
Sequence 3, Application US/08864038A  
Patent No. 6001592  
GENERAL INFORMATION:  
APPLICANT: Kunio NAKASHIMA et al.  
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: 812-5 Hirano  
STREET: Isshinoden  
CITY: Tsu-city  
STATE: Mie-prefecture  
COUNTRY: JAPAN  
ZIP: 514-01  
COMPUTER READABLE FORM: 3.50 inch, 1.44 MB storage  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Word Perfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/864, 038A  
FILING DATE: May 28, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-184459  
FILING DATE: 15-July-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: C. Bruce Hamburg  
REGISTRATION NUMBER: 22,389  
REFERENCE/DOCKET NUMBER: F-5610  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)986-2340  
TELEFAX: (212)953-7733  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Pinctada fucata  
CELL TYPE: mantle epithelial cell  
FEATURE:  
NAME/KEY: peptide  
LOCATION: from 1 to 738  
IDENTIFICATION METHOD: E (by experiment)

US-08-864-038A-3

Query Match	12.9%	Score 220.5;	DB 3;	Length 738;
Best Local Similarity	29.6%	Pred. No. 2.2e-11;		
Matches 113; Conservative	29;	Mismatches 167;	Indels 73;	Gaps 14

QY 19 AKAAEEAASASATGNAALIGDYVKNNGSAAAKGCEAASVNGIKNGIKGIYDAGKMDAKE 78  
 Db 324 AAAAAAAAAAGGGGGGGGGGGGGGGGAAAAAAAAASASASR0M8GTIDALG--D1KD 381  
 QY 79 GKLDATGAECTTNYNAGKL-FVKKRAADDGDA--DDAG-----KAAAYAAASATGMA 129  
 Db 382 -LIRSNNGSASAKASAKASAVASTKQIDIDLMDYLKDLAIGLKSSASASASASASASAGGCG 440  
 QY 130 IGDYVNGDVAKAKKGDAASVNGI-AKTIKGIYDAAEKADAKKEGKLNAAGAECTTNADAGK 188  
 Db 441 GGGNGGGGGGGGGGAGALAAALLMAAGAGGGLGCGGGGAGLAAALMAAGCGGFGGLG 500  
 QY 189 LFVNAGNVGCEACDACKAAAAVAAVSGEQLKALVIAAAMDGEKCKKADPTNPIDAA 248  
 Db 501 L-----GGLGGGSAAAAAAAAAAAGGGGRLRRRLRR0RNGG--GSAAAAAAAAAA 552  
 QY 249 IGC-----ACGNDAAAAFATMKKDDQJIAAAVYL-----RG 278  
 Db 553 GCGMGCGMGCGGFGYGLGCGFGCGGCGSSAAAAA-----AAAAGFGGGRGGRGRG 605  
 QY 279 MAKDSQ---FALKDAAAHHEGTVKNAVDIITKAAAEASASASATGSAIIGDYV----- 328  
 Db 606 RCGGDGDDNGASAVAAAAAAGSAAADY-----AAAAAAAMWYGDGADGDPFDNCGFG 661  
 QY 329 -NGNGATATKAGDASKSVNGIANG 349  
 Db 662 GNGNGCGGSGGGSGGGSGGG 683

RESULT 2  
 US-09-247-806-6  
 : Sequence 6: Application US/09247806  
 : Patent No. 6280747  
 : GENERAL INFORMATION:  
 : APPLICANT: PHILIPPE, Michel  
 : APPLICANT: GARSON, Jean-Claude  
 : APPLICANT: ARRAUDEAU, Jean-Pierre  
 : TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
 : TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
 : TITLE OF INVENTION: ANALOG  
 : FILE REFERENCE: 6388-0365-0  
 : CURRENT APPLICATION NUMBER: US/09/247,806  
 : CURRENT FILING DATE: 1999-02-11  
 : EARLIER APPLICATION NUMBER: FR 98/01614  
 : EARLIER FILING DATE: 1998-02-11  
 : NUMBER OF SEQ. ID NOS.: 14  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ. ID NO. 6  
 : LENGTH: 606  
 : TYPE: PRT  
 : ORGANISM: *Phila clavipes*  
 : US-09-247-806-6

Query Match	12.1%	Score 207.5;	DB 4;	Length 606;
Best Local Similarly	26.8%	Pred. No. 2.3e-10;		
Matches 102; Conservative	32;	Mismatches 182;	Indels 65;	Gaps 9

[illegible]

OY 136 GDVAKAKGGDAASVNGIAKGIKGIYD-----AAEKADPKBEKLVNAGMEGT 181  
 Db 349 QGAGCGAGACAAAAAGAGAGCGGTGGLGSGAGRGCGGAGAAAAAAGAGCGGTGGLGSGGA 408  
 OY 182 TNADAGKLEFVNAGNVGGEADGAKAAAVAAVSGEUIL-----KATVAAKDKGGEKO 234  
 Db 409 GGGGTGGTGGSGAGRGGLGCGAGAAAAAAGCGGGGGLGSGGAGCGAAAAAAGAGCG 468  
 OY 235 -----GKKAADRTNPIDALIGSGDNDAAAAPATMCK-----DDIIAAMVLRGMAKD 282  
 Db 469 GGTGGTGGSGGGR-----GGGAGAGAAAAAGAGAGCGGTGGLGSGAGAGCGGTGGLGSG 520  
 OY 283 -----GQFALKDAAAHEGTVKNVDTIITKAAAEASASAAATGSA-----ATGDDV 328  
 Db 521 GAGRGGLGCGCGAGAAAAAAGAGAGCGGTG--SOGAGCGAGAAAAAAGAGAGCGGTGGLGSG 579  
 OY 329 NGNGATATGSDAKSVNGIANG 349  
 Db 580 AGRGCGGAGAAAAAAGAGCG 600

```

1      RESULT      3
2      US-08-556-978B-63
3      : Sequence 63, Application US/08556978B
4      : Patent No. 6268169
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: FANNESTOCK, STEPHEN F.
8      : TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
9      : TITLE OF INVENTION: SPIDER SILK ANALOGS
10     : NUMBER OF SEQUENCES: 107
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
13     : STREET: 1007 MARKET STREET
14     : CITY: WILMINGTON
15     : STATE: DELAWARE
16     : COUNTRY: UNITED STATES OF AMERICA
17     : ZIP: 19898
18     :
19     : COMPUTER READABLE FORM:
20     : MEDIUM TYPE: DISKETTE, 3.50 INCH
21     : COMPUTER: IBM PC COMPATIBLE
22     : OPERATING SYSTEM: MICROSOFT WINDOWS 95
23     : SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
24     : CURRENT APPLICATION DATA:
25     : APPLICATION NUMBER: US/08/556,978B
26     : FILING DATE:
27     : CLASSIFICATION: 435
28     : PRIOR APPLICATION DATA:
29     : APPLICATION NUMBER: 08/077,600
30     : FILING DATE: JUNE 15, 1993
31     : ATTORNEY/AGENT INFORMATION:
32     : NAME: FLOYD, LINDA AXAMETHY
33     : REGISTRATION NUMBER: 33,692
34     : REFERENCE/DOCKET NUMBER: CR-9389-A
35     : TELECOMMUNICATION INFORMATION:
36     : TELEPHONE: 302-892-8112
37     : TELEFAX: 302-773-0164
38     : INFORMATION FOR SEQ ID NO: 63:
39     : SEQUENCE CHARACTERISTICS:
40     : LENGTH: 604 amino acids
41     : TYPE: amino acid
42     : STRANDEDNESS: unknown
43     : TOPOLOGY: unknown
44     : MOLECULE TYPE: protein
45     :
46     : US-08-556-978B-63

```

```

Query Match      11.9%  Score 2.5:  DR 4:  Length 604;
Best Local Similarity 26.6%  Pred. No. 5,2e-10;
Matches 102;  Conservative 35;  Mismatches 162;  Indels 85;  Gaps 13.

      19  AKAAEAASAAASAAATGNA-----AIGDVKVNSGAAGKGAASVNGIAKGIKGLVDAAG  72
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      247  SGGAGCGGGAAGAAAAAGGAGCGGCGTGGTGGGAGRCGGGAGAAAAAAGGAGGCGTGGTGGSGG  306
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



QY 73 KADAECKLDATGAEETTNNVNGKLFVKRAADGGDADDAAGKAAAVASATGNNAIGD 132  
 Db 307 -----GCGGGLGSGG-----AGR-----GC-----LGCGAGAAAAAAGGAGCGG 343  
 QY 133 VVNGDVAKKGGDASVNGIANGIKGIYD-----AAEKADAEKGLMAAGA 178  
 Db 344 LGSQAGGAGGAGAAAAAGAGGAGGAGGGLGSGAGRGCGGAGAAAAAAGAGAGGCGGLGS 403  
 QY 179 EETTNAADACKLFVKNAAGNVGEGADGAKAAAVAAVSGEOLL-----KAIVHAAKDGG 231  
 Db 404 QGAGGCGGGLGSGAGRGGLGCGGAGAAAAAAGAGGAGGGLGSGAGCGAGAAAAAAGG 463  
 QY 232 EKQ-----GKKAADRTNPIDAAIGAGDNDAAAFAFMKK-----DDQIAAAVLRGM 279  
 Db 464 AGCGGGLGSGAGR-----GGGAGAGAAAAAAGAGGAGGCGGLGSGAGGAGGGLG 515  
 QY 280 AKD-----GQFALKDAAAHEGTVKNAVDIIRKAAEAASASATGSA-----AIG 325  
 Db 516 GSGAGRGGLGSGAGAAAAAAGAGGAGGGLG-SGAGGAGGAGAAAAAAGAGAGGCGGLG 574  
 QY 326 DVVNGNGATKAGGDAKSVNGIAKG 349  
 Db 575 SGGAGRGCGGAGAAAAAAGAGGAGG 598

## RESULT 4

US-08-556-978B-23  
 ; Sequence 23, Application US/08556978B  
 ; Patent No. 6268169

## GENERAL INFORMATION:

APPLICANT: FAHNESTOCK, STEPHEN F.  
 TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED  
 TITLE OF INVENTION: SPIDER SILK ANALOGS  
 NUMBER OF SEQUENCES: 107  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
 STREET: 1007 MARKET STREET  
 CITY: WILMINGTON

STATE: DELAWARE  
 COUNTRY: UNITED STATES OF AMERICA

ZIP: 19898  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.50 INCH

COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: MICROSOFT WINDOWS 95

SOFTWARE: MICROSOFT WORD FOR WINDOWS 95  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/556,978B  
 FILING DATE:

CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/077,600  
 FILING DATE: JUNE 15, 1993

ATTORNEY/AGENT INFORMATION:  
 NAME: FLOYD, LINDA AXAMETHY

REGISTRATION NUMBER: 33,692  
 REFERENCE/DOCKET NUMBER: CR-9389-A

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 302-892-8112

TELEFAX: 302-773-0164  
 INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 606 amino acids

TYPE: amino acid  
 STRANDEDNESS: unknown

TOPOLOGY: unknown  
 MOLECULE TYPE: protein

US-08-556-978B-23

Query Match 11.9%; Score 203.5; DB 4; Length 606;  
 Best Local Similarity 26.5%; Pred. No. 5.2e-10;

Matches 101; Conservative 32; Mismatches 183; Indels 65; Gaps 9;  
 QY 25 AASASATGNAITGVNKNNGSAAKGEASVNGIAKTIKTIY-----DAAGKAD 75  
 Db 25 AGAAAAAAGAGGAGGGLGSGAGGAGGAGAAAAAAGAGGAGGGLGSGAGRGCGAGAAA 84  
 QY 76 AKEGKLDATGAEETTNNVNGKLFVKRAADGGDADDAAGKAAAVASATGNNAIGD 135  
 Db 85 AAAGGAGGCGGGLGSGAGGAGGGLGSGAGRGGLGSGAGGAGAAAAAAGAGGCGGLGS 144  
 QY 136 GDAVAKKGGDASVNGIANGIKGIYD-----AAEKADAEKGLMAAGAEGT 181  
 Db 145 QGAGGAGGAGAAAAAGAGGAGGAGGGLGSGAGRGCGGAGAAAAAAGAGAGGCGGLGS 204  
 QY 182 TNADACKLFVKNAAGNVGEGADGAKAAAVAAVSGEOLL-----KAIVHAAKDGG 234  
 Db 205 GCGGGLGSGAGRGGLGCGGAGAAAAAAGAGGAGGCGGLGSGAGCGAGAAAAAAGAGG 264  
 QY 235 -----GKKAADRTNPIDAAIGAGDNDAAAFAFMKK-----DDQIAAAVLRGM 282  
 Db 265 GCGGGLGSGAGR-----GGGAGAGAAAAAAGAGGAGGCGGLGSGAGGAGGGLG 316  
 QY 283 -----GQFALKDAAAHEGTVKNAVDIIRKAAEAASASATGSA-----AIGDVV 328  
 Db 317 GAGRGGLGSGAGAAAAAAGAGGAGGGLG-SGAGGAGGAGAAAAAAGAGAGGCGGLG 375  
 QY 329 NNGATKAGGDAKSVNGIAKG 349  
 Db 376 AGRGCGGAGAAAAAAGAGGAGG 396

## RESULT 5

US-09-247-806-8  
 ; Sequence 8, Application US/09247806  
 ; Patent No. 6280747

## GENERAL INFORMATION:

APPLICANT: PHILLIPE, Michel  
 APPLICANT: GARSON, Jean-Claude

APPLICANT: ARRAUDEAU, Jean-Pierre  
 TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTRACTING AT  
 TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN

FILE REFERENCE: 6388-0365-0  
 CURRENT APPLICATION NUMBER: US/09/247,806

CURRENT FILING DATE: 1999-02-11  
 EARLIER APPLICATION NUMBER: FR 98/01614

EARLIER FILING DATE: 1998-02-11  
 NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 8

LENGTH: 606  
 TYPE: PRT

ORGANISM: Nephila clavipes  
 US-09-247-806-8

Query Match 11.9%; Score 203.5; DB 4; Length 606;  
 Best Local Similarity 26.5%; Pred. No. 5.2e-10;

Matches 101; Conservative 32; Mismatches 183; Indels 65; Gaps 9;

QY 25 AASASATGNAITGVNKNNGSAAKGEASVNGIAKTIKTIY-----DAAGKAD 75  
 Db 25 AGAAAAAAGAGGAGGGLGSGAGGAGGAGAAAAAAGAGGAGGGLGSGAGRGCGAGAAA 84  
 QY 76 AKEGKLDATGAEETTNNVNGKLFVKRAADGGDADDAAGKAAAVASATGNNAIGD 135  
 Db 85 AAAGGAGGCGGGLGSGAGGAGGGLGSGAGRGGLGSGAGGAGAAAAAAGAGGCGGLGS 144  
 QY 136 GDAVAKKGGDASVNGIANGIKGIYD-----AAEKADAEKGLMAAGAEGT 181  
 Db 145 QGAGGAGGAGAAAAAGAGGAGGAGGGLGSGAGRGCGGAGAAAAAAGAGAGGCGGLGS 204  
 QY 182 TNADACKLFVKNAAGNVGEGADGAKAAAVAAVSGEOLL-----KAIVHAAKDGG 234



OY	20	KAABEASAAASATGNMAIGDV-VKNSGAAAKGE----	AASVNGIAKGIKIVDAAK	73	
Db	234	QAGAGGAGASAAAAAGAGGCGGGLGSGAGRCGEGAAAAAAGGAGCGGGLGSGGA	293		
OY	74	ADAREKGLIDATGCA--EGTTNNAGKLEFYKRAADGGBADDAKRAAAVAASA-----	123		
Db	294	GGGGVLGSGGACRRCGLGCGGCAAAAAAGAGCGGLCGAGAGCGAGAAAAAGGAGCGGY	355		
OY	124	-----ATGNMAIGDVMVNGVYAKKGGDAA-----	SVNGIAKGIKIVDAEKAKE	170	
Db	354	GGLSGCGAGRCGLGCGCAGAVAAAAAGAGCGGGLGSGCAGHCGGG-----	ACAAAAAA	409	
OY	171	GKLNAAAGAEETTADAGKLEFYKNAWNTGGEAGDAGKAAVAAVASGEOILKAIIVHAAKDG	230		
Db	410	GGAGCGRCRGGLGNGCAGC-----	GLGCGG--AGAAAAAAGGAGCGGCGGL---	GNCG	458
OY	231	GEKGGKKAADHTNPIDAAIGGAGCN-----	DAAAAFATKKKDDQITAA	273	

Query Match	11.5%;	Score 197;	DB 2;	Length 718;
Best Local Similarity	26.1%;	Pred. No. 2,4e-09;		
Matches	98;	Conservative 32;	Mismatches 164;	Indels 82; Gaps 13;
QY	20	KAAEEAASASAAVTNNAIGDV-VKNSGAAAG-GE----	AAVSNGIAKIGIYDAAGK	73
Db	234	QAGAGGACASAAAAGCAAGCGYGGIGLSQAGAGGCRACAAAAAAGACAGCGYGGIGLGGCA		293
QY	74	ADAKEGCKLDATTA--EGTTNVAAGKLPFYKRADDGDDADDAACKAAAYAAASA-----		123
Db	294	GCGYGGIGLSQAGAGGIGLGGAGGAAAAGCAAGCGELGGCGAGGAAAAGGACGAGCGGY		353
QY	124	-----ATGNNAIGDVVNGDVAAKAKGDAA-----	SVNGIAKIGIYDAEAKDAKE	170
Db	354	GGLGSGAGGGRGLGGGAGAAVAAAAAAGGAGGCGYGLGSQAGRGGGG--	AGAAAAAA	409
QY	171	GKLNAAGAGETTNAADAGKLFVNKNAQNVNGEAGADGAKAAAAAANAASGGEQILKAIIVHAARDG		230

Db 410 GGAGGGRGYGLGNOGAGR-----GGLGGOG--AGAAAAAAGGAGCGGYGL---GNOG 458  
 QY 231 GEGOGKKADRTNPIDAAIGAGDN-----DAAAFAFTMKKDDQIAAA 273  
 Db 459 AGRGGGAA-----AAAGGAGGGGYGGLGSGGAGRGCGGAGAAAAAAGAGGQEGIRG- 510  
 QY 274 MYLRGMAKDGOFALKDAAAHBGTYNNAVDI1KAAAEASASASATGSAATGDPVNGNGA 333  
 Db 511 ---QGAGCGGYGGLGSGGSGRG-----LGGGAGAAAAAAGAGGGLGGGAG 557  
 QY 334 TAKGDAXSVNGIANG 349  
 Db 558 QGAGAAAAAAGGVROG 573

## RESULT 10

US-09-034-177-3  
 ; Sequence 3, Application US/09034177  
 ; Patent No. 6127146  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/034,177  
 ; FILING DATE: HEREWITH  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0486 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-855-0555  
 ; TELEFAX: 650-845-4166  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 747 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GenBank  
 ; CLONE: GI 1174414  
 ; US-09-034-177-3

Query Match 11.5%; Score 197; DB 3; Length 747;  
 Best Local Similarity 26.1%; Pred. No. 2.5e-09;  
 Matches 98; Conservative 32; Mismatches 164; Indels 82; Gaps 13;

QY 20 KAEFEASASASATGMAATGADY-VKNSGAAGGE-----AASVNGIANGIGYDPAACK 73  
 Db 234 QGAGGAGASASAAAGAGCGGYGGLGSGAGRGAGAGAAAAAAGAGGAGCGGYGLGGGAG 293

QY 74 ADAKEGLDPTATGA--EGTTNNAGKLFYVRKADYGDADDAGKAAAVASA----- 123  
 Db 294 GGGGYGGLGSGGAGRGGLGGGAGGAGAAAGAGGAGGGLGGGAGGAGAGAGAAAAAGAGGCGY 353  
 QY 124 -----ATGNAATGDPVNGDVAKAKGDDA-----SYNGIAKGIKGYDAAEKDAKE 170  
 Db 354 GGLGSGAGRGGLGSGGAGACAVAAAAAGAGGAGGCGGYGGLGSGAGRGCGG---AGAAAAA 409  
 QY 171 GKLNAGAGETTNADAGKLFYKNAAGVGEAPADGKAAAVAAVAVSGEQLKATVHAANKG 230  
 Db 410 GGAGGGRGYGLGNOGAGR-----GGLGGOG--AGAAAAAAGGAGCGGYGL---GNOG 458  
 QY 231 GEGOGKKADRTNPIDAAIGAGDN-----DAAAFAFTMKKDDQIAAA 273  
 Db 459 AGRGGGAA-----AAAGGAGGGGYGGLGSGGAGRGCGGAGAAAAAAGAGGQEGIRG- 510  
 QY 274 MYLRGMAKDGOFALKDAAAHBGTYNNAVDI1KAAAEASASASATGSAATGDPVNGNGA 333  
 Db 511 ---QGAGCGGYGGLGSGGSGRG-----LGGGAGAAAAAAGAGGGLGGGAG 557  
 QY 334 TAKGDAXSVNGIANG 349  
 Db 558 QGAGAAAAAAGGVROG 573

## RESULT 11

US-08-209-747-2  
 ; Sequence 2, Application US/08209747  
 ; Patent No. 5733771  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lewis, Randolph V.  
 ; APPLICANT: Colgin, Mark  
 ; TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider  
 ; NUMBER OF SEQUENCES: 56  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
 ; STREET: P.O. Box 747  
 ; CITY: Falls Church  
 ; STATE: Virginia  
 ; COUNTRY: USA  
 ; ZIP: 22040-3487  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/209,747  
 ; FILING DATE: 14-MAR-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Murphy Jr., Gerald M.  
 ; REGISTRATION NUMBER: 28,977  
 ; REFERENCE/DOCKET NUMBER: 1447-104P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-205-8000  
 ; TELEFAX: 703-205-8050  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 832 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: N. clavipes  
 ; TISSUE TYPE: minor ampullate gland  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..309



```

? REFERENCE/DOCKET NUMBER: CR-9289-A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 302-892-8112
? TELEFAX: 302-773-0184
? INFORMATION FOR SEQ ID NO: 21:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 606 amino acids
? TYPE: amino acid
? STRANDEDNESS: unknown
? TOPOLOGY: unknown
? MOLECULE type: protein
? IS-ORF-556-978B-21

```

Query Match	11.48	Score 194.5	DB 4	Length 606
Best Local Similarity	26.48	Pred. No. 3.1e-09		
Matches 96	Conservative 31	Mismatches 176	Indels 61	Gaps 12

[illegible]

T 14  
 US-09-247-806-4  
 : Sequence 4, Application US/09247806  
 : Patent No. 6280747  
 : GENERAL INFORMATION:  
 : APPLICANT: PHILIPPE, Michel  
 : APPLICANT: GARSON, Jean-Claude  
 : APPLICANT: ARAUDEAU, Jean-Pierre  
 : TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
 : TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
 : TITLE OF INVENTION: ANALOG  
 : FILE REFERENCE: 6388-0365-0  
 : CURRENT APPLICATION NUMBER: US/09/247,806  
 : CURRENT FILING DATE: 1999-02-11  
 : EARLIER APPLICATION NUMBER: FR 98/01614  
 : EARLIER FILING DATE: 1998-02-11  
 : NUMBER OF SEQ ID NOS: 14  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 4  
 : LENGTH: 606  
 : TYPE: PRT  
 : ORGANISM: Nephila clavipes  
 : US-09-247-806-4

Query Match	11.48; Score 194.5; DB 4; Length 606;
-------------	---------------------------------------

Best Local Similarity 26.4%; Pred. No. 3.1e-09;  
Matches 96; Conservative 31; Mismatches 176; Indels 61; Gaps 12

Oy	25	AASAAATGTAATIGDYYKXNGAANKGGEPAASVNGTAKGTIGYIDVADAKMDAKEGKIDAT	84
Db	42	AGAAAAAAGGAGGGGGLGSGCAGCAGAAAAAAGAGCGGGTGLGSGCAGCGGTGLGSGO	101
Oy	85	GA-----EGTTNNAGKLFVKRAADDG-----TADDAKRA-----AAAVASATGNA	129
Db	102	GAGRGCGCAGAAAAAG-----GAGGCGTGTGLTSGCAGRGGLGCGGAGAAAAAAGGAG	155
Oy	130	IGDYVNGDVAKAKGDAASVNGTAKGTIKGYIDAAEKDAKEGKILNAGA-----EGTTNADAK	188
Db	156	QGGGLSGCAGCAGAAAAAAGCAGCGGTGGLGSGCAGCGGTGGLGSGCAGRCGCGAGAAA	215
Oy	189	LEFVNAGNVG-----GEAGDAGKMAAANVAASGFGILKALYIHAAKDDGEGKGK	237
Db	216	AAAGCAGGGGTGGLGSGCAGRGGLGCGGAGAAAAAAGC-----ACGGGLGSGG--	264
Oy	238	AADPTNPIDAAIGGAGGNDAAAAAATATKKDDQLAAAVYLGRMKMDGQFALUKDAAAA----	293
Db	265	AGCGAGAAAAAAGGAGGG--GYGGGLSGGAGGGGTGGLGSGCAGRGCGGAGAAAAAAGGAG	323
Oy	294	-----HBCTVKNAYDIITKAAAEPAASAAATGSAIIGDYVNGCATKKGGDASVNG	345
Db	324	QGGGTGGLGSGCAGRGGLGCGGAGAAAAAAGGA--GCGGLGSGCAGGA-----GAAAAAAGG	379
Oy	346	IAKG 349	
Db	380	AGCG 383	

```

15 RESULT
16 US-07-941-523-24
17 ; Sequence 24, Application US/07941523
18 ; Patent No. 5571718
19 ; GENERAL INFORMATION:
20 ; APPLICANT: Dunn, John J
21 ; APPLICANT: Barbour, Alan G
22 ; TITLE OF INVENTION: Cloning and Expression of Borrelia
23 ; TITLE OF INVENTION: Lipoproteins
24 ; NUMBER OF SEQUENCES: 24
25 ; CORRESPONDENCE ADDRESS:
26 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
27 ; STREET: Two Milltia Drive
28 ; CITY: Lexington
29 ; STATE: Massachusetts
30 ; COUNTRY: U.S.A.
31 ; ZIP: 01730
32 ; COMPUTER READABLE FORM:
33 ; MEDIUM TYPE: Floppy disk
34 ; COMPUTER: IBM PC compatible
35 ; OPERATING SYSTEM: PC-DOS/MS-DOS
36 ; SOFTWARE: Patent Release #1.0, Version #1.25
37 ; CURRENT APPLICATION DATA:
38 ; APPLICATION NUMBER: US/07/941,523
39 ; FILING DATE: 19920908
40 ; CLASSIFICATION: 435
41 ; ATTORNEY/AGENT INFORMATION:
42 ; NAME: Granahan, Patricia
43 ; REGISTRATION NUMBER: 32,227
44 ; REFERENCE/DOCKET NUMBER: BN190-01A
45 ; TELECOMMUNICATION INFORMATION:
46 ; TELEPHONE: (617) 861-6240
47 ; TELEFAX: (617) 861-9540
48 ; INFORMATION FOR SEQ ID NO: 24:
49 ; SEQUENCE CHARACTERISTICS:
50 ; LENGTH: 344 amino acids
51 ; TYPE: AMINO ACID
52 ; STRANDEDNESS: single
53 ; TOPOLOGY: linear
54 ; MOLECULE TYPE: protein
55 ; US-07-941-523-24

```

Query Match 11.38; Score 193; DB 1; Length 344;  
Best Local Similarity 25.68; Pred. No. 2.1e-09;  
Matches 91; Conservative 50; Mismatches 149; Indels 66; Gaps 14;

OY 18 MAKAAEEAASAAATGNAAGDVVANSNGAAGGAASVNGIAKGIKIVDAAGKADAK 77  
DB 1 MAGQPEACKTGVSGVNGNLGNSLMLGRSAENAFYAFI---ELVSDVLGFTAKSDTT 56  
OY 78 E---GKLDATGAE-GTTVNVNAGKLFVKRAADGDGADDAKAAAVAAASAA----- 124  
DB 57 KQEVGGYFNSLGAKEASNDLEQVAVK--AETGVDSKSDSKNPIREAVNEAKEVLGTLK 114  
OY 125 ---TGNAAGD---VYNGDVAKAKAGDAA--SYNGIAKGIKIVDAAEKADAKEGLNAA 176  
DB 115 GYVESLGTIGDSNPVGYANNAAGSGTTAADDELKRAFKALQETIVKAAATDAGVKALKIGAT 174  
OY 177 GAE---GTTNADAGKLFVKNAGNVGGEAGDAGKAAAVAAVSGEOTLKATVHAAKDGGEK 233  
DB 175 TLDANGGADNKEGAKILATISGNPA--AADVAKAAAILSSVSGEEMLSIV-----K 224  
OY 234 QGKKAADRTNPIDAAIGAGD-NDAAAFATMKKDDQIAAAVLRGMKXDGDFALKDAAA 292  
DB 225 SGEN-----DAQLAAADGNTSAISFAKGGSDAHLGANTPKAAAVAGIALR----- 272  
OY 293 AHEGYKNAVDITKAAAEASAAATGSAAGDVVNGATAGKGDGAKSVNGCIATK 348  
DB 273 -----SLVKTGKLAAAGAADNATG---GEYGVGVAAANKLLRAVEDVIK 315

Search completed: January 10, 2002, 14:03:56  
Job time: 90 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 184.82 Seconds  
(without alignments)  
524.306 Million cell updates/sec

Title: US-09-445-803-2

Perfect score: 1708

Sequence: 1 KNNDDHNNHGTAKNAVDMAK.....GNCATKAGDCAKSVNGLAKG 349

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

3148936

number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents: AA Main:\*

1: /cgn2\_6/ptodata/2/paa/PCrus.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US081.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US080.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US082.COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083.COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084.COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085.COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087.COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088.COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089.COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090.COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091.COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092.COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093.COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094.COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095.COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096.COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097.COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098.COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099.COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US60.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1708	100.0	349	18	US-09-445-803-2
2	1702	99.6	349	17	US-09-300-971A-9
3	532	31.1	323	17	US-09-300-971A-11
4	480	28.1	189	17	US-09-300-971A-10
5	437.5	25.6	356	1	PCT-US97-02952-2
6	437.5	25.6	356	15	US-09-125-619-2
7	437.5	25.6	356	15	US-09-125-619-13
8	417.5	24.4	208	15	US-09-125-619-47
9	416	24.4	211	15	US-09-125-619-37

10	412.5	24.2	214	15	US-09-125-619-34	Sequence 34, Appl
11	409.5	24.0	212	15 <td>US-09-125-619-31</td> <td>Sequence 31, Appl</td>	US-09-125-619-31	Sequence 31, Appl
12	409.5	24.0	212	15 <td>US-09-125-619-43</td> <td>Sequence 43, Appl</td>	US-09-125-619-43	Sequence 43, Appl
13	409.5	24.0	216	15 <td>US-09-125-619-38</td> <td>Sequence 38, Appl</td>	US-09-125-619-38	Sequence 38, Appl
14	408	23.9	215	15 <td>US-09-125-619-36</td> <td>Sequence 36, Appl</td>	US-09-125-619-36	Sequence 36, Appl
15	405.5	23.7	214	15 <td>US-09-125-619-35</td> <td>Sequence 35, Appl</td>	US-09-125-619-35	Sequence 35, Appl
16	405	23.7	209	15 <td>US-09-125-619-33</td> <td>Sequence 33, Appl</td>	US-09-125-619-33	Sequence 33, Appl
17	405	23.7	213	15 <td>US-09-125-619-39</td> <td>Sequence 39, Appl</td>	US-09-125-619-39	Sequence 39, Appl
18	404	23.7	211	15 <td>US-09-125-619-40</td> <td>Sequence 40, Appl</td>	US-09-125-619-40	Sequence 40, Appl
19	403	23.6	209	15 <td>US-09-125-619-42</td> <td>Sequence 42, Appl</td>	US-09-125-619-42	Sequence 42, Appl
20	398.5	23.3	216	15 <td>US-09-125-619-46</td> <td>Sequence 46, Appl</td>	US-09-125-619-46	Sequence 46, Appl
21	398	23.3	209	15 <td>US-09-125-619-45</td> <td>Sequence 45, Appl</td>	US-09-125-619-45	Sequence 45, Appl
22	397.5	23.3	212	15 <td>US-09-125-619-48</td> <td>Sequence 48, Appl</td>	US-09-125-619-48	Sequence 48, Appl
23	392.5	23.0	212	15 <td>US-09-125-619-32</td> <td>Sequence 32, Appl</td>	US-09-125-619-32	Sequence 32, Appl
24	390	22.8	189	15 <td>US-09-125-619-40</td> <td>Sequence 20, Appl</td>	US-09-125-619-40	Sequence 20, Appl
25	389.5	22.8	189	15 <td>US-09-125-619-44</td> <td>Sequence 44, Appl</td>	US-09-125-619-44	Sequence 44, Appl
26	387	22.7	187	15 <td>US-09-125-619-23</td> <td>Sequence 23, Appl</td>	US-09-125-619-23	Sequence 23, Appl
27	384	22.5	190	15 <td>US-09-125-619-15</td> <td>Sequence 15, Appl</td>	US-09-125-619-15	Sequence 15, Appl
28	382.5	22.4	212	15 <td>US-09-125-619-41</td> <td>Sequence 41, Appl</td>	US-09-125-619-41	Sequence 41, Appl
29	379	22.2	190	15 <td>US-09-125-619-19</td> <td>Sequence 19, Appl</td>	US-09-125-619-19	Sequence 19, Appl
30	377.5	22.1	195	15 <td>US-09-125-619-17</td> <td>Sequence 17, Appl</td>	US-09-125-619-17	Sequence 17, Appl
31	373	21.8	197	15 <td>US-09-125-619-42</td> <td>Sequence 25, Appl</td>	US-09-125-619-42	Sequence 25, Appl
32	364.5	21.3	169	15 <td>US-09-125-619-25</td> <td>Sequence 22, Appl</td>	US-09-125-619-25	Sequence 22, Appl
33	363.5	21.3	194	15 <td>US-09-125-619-28</td> <td>Sequence 28, Appl</td>	US-09-125-619-28	Sequence 28, Appl
34	362.5	21.2	169	15 <td>US-09-125-619-24</td> <td>Sequence 24, Appl</td>	US-09-125-619-24	Sequence 24, Appl
35	361.5	21.2	190	15 <td>US-09-125-619-29</td> <td>Sequence 29, Appl</td>	US-09-125-619-29	Sequence 29, Appl
36	359	21.0	179	15 <td>US-09-125-619-18</td> <td>Sequence 18, Appl</td>	US-09-125-619-18	Sequence 18, Appl
37	359	21.0	190	15 <td>US-09-125-619-27</td> <td>Sequence 27, Appl</td>	US-09-125-619-27	Sequence 27, Appl
38	345.5	20.2	189	15 <td>US-09-125-619-16</td> <td>Sequence 26, Appl</td>	US-09-125-619-16	Sequence 26, Appl
39	341	20.0	169	15 <td>US-09-125-619-16</td> <td>Sequence 16, Appl</td>	US-09-125-619-16	Sequence 16, Appl
40	319	18.7	158	15 <td>US-09-125-619-21</td> <td>Sequence 21, Appl</td>	US-09-125-619-21	Sequence 21, Appl
41	266.5	15.6	339	15 <td>US-09-125-619-14</td> <td>Sequence 14, Appl</td>	US-09-125-619-14	Sequence 14, Appl
42	242	14.2	157	15 <td>US-09-125-619-30</td> <td>Sequence 30, Appl</td>	US-09-125-619-30	Sequence 30, Appl
43	226.5	13.3	456	24	US-60-191-637-10600	Sequence 20600, A
44	226.5	13.3	456	24	US-60-191-661-16220	Sequence 16220, A
45	214	12.5	681	1	PCT-US00-30086-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-445-803-2  
Sequence 2, Application US/09445803  
GENERAL INFORMATION:  
APPLICANT: Adminis. of Tulane Educational, Fund  
Philip P. Mario T.  
TITLE OF INVENTION: Surface Antigens and Proteins Useful in  
Compositions for the Diagnosis and Prevention of Lyme  
Disease  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESS: Howson and Howson  
STREET: Spring House Corporate Cntr., P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/445,803  
FILING DATE: 13-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/051,271  
FILING DATE: 30-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215

```

? REFERENCE/POCKET NUMBER: TUL2APCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-540-9200
? TELEFAX: 215-540-5818
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 349 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? SEQUENCE DESCRIPTION: SEQ ID NO: 2:
? JS-09-445-803-2

```

Query Match	100.0%	Score 1708;	DB 18;	Length 349;
Best Local Similarity	100.0%	Pred. No. 9.7e-137;		
Matches 349;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	61	AKGIKGIYDADGKDAKDEKIDATGAGCTTIVNNGKLFYVKRADDDGDADAGKAAAVA	120
Dd	61	AKGIKGIYDADGKDAKDEKIDATGAGCTTIVNNGKLFYVKRAADDGDADADAGKAAAVA	120
QY	121	ASAATGNAAGIDVYNGDVAKAKAGDAAVSVNGTAKIGIYDAAEKADAKEGKLTAAAGC	180
Dd	121	ASAATGNAAGIDVYNGDVAKAKAGDAAVSVNGTAKIGIYDAAEKADAKEGKLTAAAGC	180
QY	181	TTNNDAGGLFVKNNGNNGGEGKDGKAAAVAAVSGDQILKAIYHAAKDDGEGKGGKRAAD	240
Dd	181	TTNNDAGGLFVKNNGNNGGEGKDGKAAAVAAVSGDQILKAIYHAAKDDGEGKGGKRAAD	240
QY	241	RTNPIDAAGAGGNDNAAAAAATATKCKDQDILAAAVYLRGMAKDDGOLFATDAAAHEGTVKN	300
Dd	241	RTNPIDAAGAGGNDNAAAAAATATKCKDQDILAAAVYLRGMAKDDGOLFATDAAAHEGTVKN	300
QY	301	AVDIIKAAAEAAASAATGSAATGDIYVNGNGATAKGDAAKSVNGTAKG 349	
Dd	301	AVDIIKAAAEAAASAATGSAATGDIYVNGNGATAKGDAAKSVNGTAKG 349	

```

RESULT 2
US-09-300-971A-9
Sequence 9, Application US/09300971A
GENERAL INFORMATION:
APPLICANT: Philipp, Mario T.
TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
FILE REFERENCE: TUL305A
CURRENT APPLICATION NUMBER: US/09/300,971A
CURRENT FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 349
TYPE: PRP
ORGANISM: Borrelia garlhi
US-09-300-971A-9

```

Query Match	99.68;	Score 1702;	DB 17;	Length 349;
Best Local Similarity	99.78;	Pred. No. 3.2e-136;		
Matches 348; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	QY	Db	QY
1	1	1	61
KNNDHNDHKGTVYKAAVDMAKAAEEAASAAATGNGMA	KNNDHNDHKGTVYKAAVDMAKAAEEAASAAATGNGMA	KNNDHNDHKGTVYKAAVDMAKAAEEAASAAATGNGMA	AKGKIGKIVDAAGKADAEKGLDITGAGCTTNVYNAKGLFVYKRAADDGADADAGKAAAAYA
IGGVYVNSGCAAAAGGGAASYNGI	IGGVYVNSGCAAAAGGGAASYNGI	IGGVYVNSGCAAAAGGGAASYNGI	120
60	60	60	
1	1	1	61
KNNDHNDHKGTVYKAAVDMAKAAEEAASAAATGNGMA	KNNDHNDHKGTVYKAAVDMAKAAEEAASAAATGNGMA	KNNDHNDHKGTVYKAAVDMAKAAEEAASAAATGNGMA	AKGKIGKIVDAAGKADAEKGLDITGAGCTTNVYNAKGLFVYKRAADDGADADAGKAAAAYA
IGGVYVNSGCAAAAGGGAASYNGI	IGGVYVNSGCAAAAGGGAASYNGI	IGGVYVNSGCAAAAGGGAASYNGI	120
60	60	60	

Oy	121	ASATGNAITGDVYNGDPAKAKGGDAASVNGTAKGIGCIYDAAEKKADAKEGKITLINAAGAE	180
Dd	121	ASAATGNAAITGDVYNGDPAKAKGGDAASVNGTAKGIGIYDAAEKKADAKEGKITLINAAGAE	180
Oy	181	TTNADACKLEFYKNAGNGVGEAGEADGAKAAAANVAASVEQILTKALITYHNAKDCEGCKOGKKAAD	240
Dd	181	TTNADACKLEFYKNAGNGVGEAGEADGAKAAAANVAASVEQILTKALITYHNAKDDGEGCKOGKKAAD	240
Oy	241	RTNPIDAAITCGAGNDNDAAAAFATYKKDDQIAAAMVLRGMMAKDCQFALKDAADAAAHBEGTVKN	300
Dd	241	ATNPIDAAITCGAGNDNDAAAAFATYKKDDQIAAAMVLRGMMAKDCQFALKDAADAAAHBEGTVKN	300
Oy	301	AVDTITKAAAEASAASTGSAAITGDVYNGGATAKGDAAKSVMGTAKG	349
Dd	301	AVDTITKAAAEASAASTGSAAITGDVYNGGATAKGDAAKSVMGTAKG	349

```

RESULT      3
US-09-300-971A-11
: Sequence 11, Application US/09300971A
: GENERAL INFORMATION:
: APPLICANT: Philipp, Mario T.
: APPLICANT: Liang, Fang Ting
: TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
: FILE REFERENCE: TUI3USA
: CURRENT APPLICATION NUMBER: US/09/300,971A
: CURRENT FILING DATE: 1999-04-28
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 11
: LENGTH: 323
: TYPE: PRN
: ORGANISM: Borrelia burgdorferi
: US-09-300-971A-11

```

Query Match	31.18;	Score 532;	DB 17;	Length 323;
Best Local Similarity	42.98;	Pred. No. 9.5e-37;		
Matches 147;	Conservative 42;	Mismatches 110;	Indels 44;	Gaps 11

```

QY 31 AANAALIGDVK--NSGAANKGEAAVSVNGIAKIGINGYDAAGKADAKKGGKIDATGAEC 88
Db 1 AVSGEOLISALVIAKAAAGAAADQDEKP--GDAK -NPILAAIGKNVADGADFGCGMK 55
QY 89 TTNNVA-----GKLFVKRAADGDGADDDGKAAAV-----AASAAATGNA 128
Db 56 DDOIAAAILRGMAKDKGFVAK--DEKGAABEAIKASELLOLVKVAVYTAGCASSGTA 113
QY 129 AIGDVNGDVAKAKGGDAASVNGIAKIGIKIVDAEKADAKBEKLLNAAAGETTNADAG 188
Db 114 AIGEYVDN--AAKAADKDSVTGIANKIKETIVEAAJCS---KLKVAAKKGNNNGAG 166
QY 189 LFVYNAGVGEACDAGKAAAAAANAASVEOILKAIYHAAADGEGKCKKAADTNPIDAA 248
Db 167 LFGRAGANAMHDSFAASKAGAVSAVSGEOLISALVIAKGEACDQDEBKKEFEAKNPILAA 226
QY 249 ICGAGDDAANAAPFT--MKKDDOIAAMVLERMAKXQCFAPKAAAAAHECTVANAADILK 306
Db 227 I---GDDGDGEFMDQMKKKDDOIAAIALRGMAKDKGFVAKGGEKAECAIGVSELDD 283
QY 307 AAAEASASAAATGSAIIGDVNGCATATGAGDKASVNGIAK 349
Db 284 LVKAVKTAEGASSGTAIGEV--ADAAVVDKASVTGIANK 323

```

```

RESULT      4
US-09-300-971A-10
: Sequence 10, Application US/09300971A
: GENERAL INFORMATION:
: APPLICANT: Phillip, Mario T.
: TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease

```

```
; FILE REFERENCE: TUL3USA
; CURRENT APPLICATION NUMBER: US/09/300,971A
; CURRENT FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 10
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-300-971A-10
```

```
Query Match          28.1%; Score 480; DB 17; Length 189;
Best Local Similarity 64.6%; Pred. No. 1.2e-32;
Matches 115; Conservative 16; Mismatches 39; Indels 8; Gaps 5;
```

```
OY 117 AAVASAATGN-AAIGDVV--NGDVAKAKGDAASVNGIAKIGIYDAEAKADAEK 172
DB 1 ATEAATAASGDKEMIGKVYKTNAGAAAKGGEKSVNGIAGIKIGIYDAEAKA-CKECK 59
173 LMAAGBETTADAGKLFYKKNAGNNGEAGDAGKAAAVAAVSGEOLIKAIYHAAKDGSE 232
DB 60 LEAAGDGKDKACAGKLFKKNNAANGGGAAGAAAEKAAAVASVSGKQIKAIYDAA--GKE 117
OY 233 KGC-KKADRTNPIDAAIGAGDNDAAAFAFMKKDDQIAAAVLRGMKDDGFALKD 289
DB 118 EKGVAADVKEATNPIDAAIGSTGDNDAAFQDEMKKDDQIAAAVLRGMKDDGFALKD 175
```

```
RESULT 5
PCT-US97-02952-2
; Sequence 2, Application PC/TUS9702952
; GENERAL INFORMATION:
```

```
; APPLICANT:
; TITLE OF INVENTION: VMP-Like Sequences of Pathogenic Borrelia
; NUMBER OF SEQUENCES: 8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/02952
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US97-02952-2
```

```
Query Match          25.6%; Score 437.5; DB 1; Length 356;
Best Local Similarity 46.9%; Pred. No. 1.2e-28;
Matches 127; Conservative 26; Mismatches 87; Indels 31; Gaps 11;
```

```
OY 79 GKLDTAGAGTTNNVNAAGKLFVRRADDGDDADGAKAAAVASAATGNAAGDVVNGDV 138
DB 103 GKPDSTGSVGT-----VEGAIKEVSELDKLVKAVKTBGASSGTAAIGEVV-ADA 153
OY 139 AKAKGDAASVNGIAKIGIYDAEAKADAEKGLNA-AGAEGTTNADAGKLFYKKNAGV 197
DB 154 DAAKADKASVNGIAKIGIYDAEAGGSE---KLKAVAAAGENNKGAGKLFGRKAGAA 209
OY 198 GGEAGDAGKAAAVAAVSGEOLIKAIYHAAKDGEGKGGKADRTNPIDAAIGAGDND 257
DB 210 HGDSEAKSAAAGAVASVSGEOLISATVTA--DAEDGKPPKAPPIAAI---GDKDG 265
OY 258 AAFAA--TKKKDDQIAAAVLRGMKDDGFALKDAAAHEGTVKKNVNDITKAAAEAA--S 313
DB 266 GAFFGODEMKKDDQIAAAVLRGMKDDGFALKD---GEKEKAGAIKGAESAVRK 319
```

```
OY 314 AASAATGSAIGDVVNGCATAKGDAKSVN 344
DB 320 VLGAITG--LIGDAVS--SGLRKVGDSVKAAS 347
```

```
RESULT 6
US-09-125-619-2
; Sequence 2, Application US/09125619
; GENERAL INFORMATION:
```

```
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-Like SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: USH:234
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-125-619-2
```

```
Query Match          25.6%; Score 437.5; DB 15; Length 356;
Best Local Similarity 46.9%; Pred. No. 1.2e-28;
Matches 127; Conservative 26; Mismatches 87; Indels 31; Gaps 11;
```

```
OY 79 GKLDTAGAGTTNNVNAAGKLFVRRADDGDDADGAKAAAVASAATGNAAGDVVNGDV 138
DB 103 GKPDSTGSVGT-----VEGAIKEVSELDKLVKAVKTBGASSGTAAIGEVV-ADA 153
OY 139 AKAKGDAASVNGIAKIGIYDAEAKADAEKGLNA-AGAEGTTNADAGKLFYKKNAGV 197
DB 154 DAAKADKASVNGIAKIGIYDAEAGGSE---KLKAVAAAGENNKGAGKLFGRKAGAA 209
OY 198 GGEAGDAGKAAAVAAVSGEOLIKAIYHAAKDGEGKGGKADRTNPIDAAIGAGDND 257
DB 210 HGDSEAKSAAAGAVASVSGEOLISATVTA--DAEDGKPPKAPPIAAI---GDKDG 265
OY 258 AAFAA--TKKKDDQIAAAVLRGMKDDGFALKDAAAHEGTVKKNVNDITKAAAEAA--S 313
DB 266 GAFFGODEMKKDDQIAAAVLRGMKDDGFALKD---GEKEKAGAIKGAESAVRK 319
OY 314 AASAATGSAIGDVVNGCATAKGDAKSVN 344
DB 320 VLGAITG--LIGDAVS--SGLRKVGDSVKAAS 347
```

```
RESULT 7
US-09-125-619-13
; Sequence 13, Application US/09125619
; GENERAL INFORMATION:
```













;; PRIOR FILING DATE: 1999-02-11  
;; PRIOR APPLICATION NUMBER: FR 98/01614  
;; PRIOR FILING DATE: 1998-02-11  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 651  
;; TYPE: PRT  
;; ORGANISM: Nephila clavipes  
US-09-861-597-1

Query Match 11.5%; Score 197; DB 5; Length 651;  
Best Local Similarity 26.1%; Pred. No. 0.00024;  
Matches 98; Conservative 32; Mismatches 164; Indels 82; Gaps 13;

QY 20 KAEERASASAAATGNAIGDV-VKNSGAARKGE-----AASVNGIAKIGIYDAACK 73  
DB 234 QGAGGAGAGASAAAAGAGCGGGLGSGGAGGEGAGAAAAAGAGAGCGGGLGCGGA 293  
C 74 AAKKCKIDATCA--EGTTNVNAGKLFYKRAADDGDDADACKAAAAVAASA----- 123  
DB 294 GGGYGGGLSGGAGRGGLGCGGAGAAAAGAGGGLGCGGAGGAAAAAGAGGCGGY 353  
QY 124 -----ATGNAIGDVNVDVAKAKGDAA-----SVNGIAKIGIYDAAEKADAKE 170  
DB 354 GGLSGGAGRGGLGCGGAGAVAAAAAGAGCGGYGGLSGGAGRGCG-----AGAAAAA 409  
QY 171 GKLNAAGEGTTNADAGKLFYNAGNVGEGADAGKAAAAVAASGEOLKAIYHAAKDG 230  
DB 410 GGAGGGRYGGGLGNOGAGR-----GGLGGG--AGAAAAAGAGGCGGYGL--GNGG 458  
QY 231 GEKQKKAADRTNPIDAIIGGAGDN-----DAAAFYTMKKDDOIAAA 273  
DB 459 ACRGCGGAA-----AAAGGAGCGGYGGLGSGGAGRGCGGAGAAAAAVGAGGEGIRG- 510  
QY 274 NLRGMAKKGDFALKDAAAHGCTVKNNAVDTIKAAEAASASAAATGSAIGDVNNGCA 333  
DB 511 ---QGAGGCGYGLSGSGSGRG-----LGGGAGAAAAAGAGGCGGLGCGGAG 557  
QY 334 TAKGDAKSVNGIANG 349  
DB 558 QGAGAAAAAGGVRGG 573

RESULT 5  
US-09-861-597-4  
;; Sequence 4, Application US/09861597  
;; GENERAL INFORMATION:  
;; APPLICANT: PHILIPPE, Michel  
;; APPLICANT: GARSON, Jean-Claude  
;; APPLICANT: ARRAUDEAU, Jean-Pierre  
;; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
;; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
;; FILE REFERENCE: 6388-0365-0  
;; CURRENT APPLICATION NUMBER: US/09/861,597  
;; PRIOR FILING DATE: 2001-05-22  
;; PRIOR APPLICATION NUMBER: 09/247,806  
;; PRIOR FILING DATE: 1999-02-11  
;; PRIOR APPLICATION NUMBER: FR 98/01614  
;; PRIOR FILING DATE: 1998-02-11  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 606  
;; TYPE: PRT  
;; ORGANISM: Nephila clavipes  
US-09-861-597-4

Query Match 11.4%; Score 194.5; DB 5; Length 606;  
Best Local Similarity 26.4%; Pred. No. 0.00031;

Matches 96; Conservative 31; Mismatches 176; Indels 61; Gaps 12;

QY 25 AASASAAATGNAIGDVYNNKSGAAAKGGEAASVNGIAKIGIYDAAGKADKEKLDAT 84  
DB 42 AGAAAAAGAGAGGGLGSGGAGCGGAGAAAAAGAGGCGGYGGLSGGAGCGGYGGLGSGO 101  
QY 85 GA-----EGTTNVNAGKLFYKRAADDG-----DADDACKA-----AAVVAASATGNA 129  
DB 102 GAGRGCGGAGAAAAAG-----GAGCGGYGGLGSGGAGRGGLGCGGAGAAAAAGAGAG 155  
QY 130 IDDVYNDVAKAKGDPAASVNGIAKIGIYDAAEKADAKEKLNAGA--EGTTNADACK 188  
DB 156 GCGGLGSGGAGGAGAAAAAGAGCGGYGGLGSGGAGCGGYGGLGSGGAGRGCGGAGAAA 215  
QY 189 LFYKNAANGV-----GEAGDAGKAAAAVAASGEOLKAIYHAAKDGEGEKGK 237  
DB 216 AAGGAGCGGYGGLGSGGAGRGGLGCGGAGAAAAAG-----AGCGGLGSGG-- 264  
QY 238 AADRTNPIDAIIGGAGDNDAAFATMKKDDOIAAMVLRGMAKDGDFALKDPAAA----- 293  
DB 265 AGCGAGAAAAAGAGAGCG--GYGGLGSGGAGCGGYGGLGSGGAGRGCGGAGAAAAAGAG 323  
QY 294 -----HEGTVKNNAVDTIKAAEAASASATGSAIGDVNNGNATRAKGDASKVNG 345  
DB 324 GGGYGGGLSGGAGRGGLGCGGAGAAAAAGGA--GGGGLGSGGAGCGA---GAAAAAGG 379  
QY 346 IAKG 349  
DB 380 AGCG 383

RESULT 6  
US-09-971-536-70  
;; Sequence 70, Application US/09971536  
;; GENERAL INFORMATION:  
;; APPLICANT: Glenn, Mathew  
;; APPLICANT: Havukala, Ilkka  
;; APPLICANT: Bloksberg, Leonard  
;; APPLICANT: Lubbers, Mark  
;; APPLICANT: Dekker, James  
;; APPLICANT: Christensson, Anna  
;; APPLICANT: Holland, Ross  
;; APPLICANT: O'Toole, Paul  
;; APPLICANT: Reid, Julian  
;; APPLICANT: Coolbear, Timothy  
;; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth  
;; FILE REFERENCE: 1043C2  
;; CURRENT APPLICATION NUMBER: US/09/971,536  
;; PRIOR FILING DATE: 2001-10-02  
;; PRIOR APPLICATION NUMBER: U.S. No. 09/634,238  
;; PRIOR FILING DATE: 2000-08-08  
;; PRIOR APPLICATION NUMBER: U.S. No. 09/724,623  
;; PRIOR FILING DATE: 2000-11-28  
;; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
;; PRIOR FILING DATE: 2001-08-08  
;; NUMBER OF SEQ ID NOS: 83  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 70  
;; LENGTH: 1879  
;; TYPE: PRT  
;; ORGANISM: Lactobacillus rhamnosus  
US-09-971-536-70

Query Match 10.7%; Score 182; DB 5; Length 1879;  
Best Local Similarity 26.6%; Pred. No. 0.0049;  
Matches 110; Conservative 35; Mismatches 154; Indels 114; Gaps 19;

QY 4 DHD-----NHKGTVKNNAV-DMAKAEASASAAATG--NAAIGDVYNNKSGAAKGEAAS 56  
DB 1252 DHDATLTANEKASOKQAVYDEATTKAKKAIDAQDAVDAQATDGTGKAIADQNHSGALD 1311

```

Qy 57 VNGIAKGICIVDAAGKADAKEGKLDATGAEETTNNACKLFVK--RAADGGDADDAK 114
Db 1312 DR-----KADAKO-VIDAEMAKVTAAIDODNTLTAKOKAGOVATEADK 1356
Qy 115 AAAAASAATGNAAGIDVYNGDVAKKAGGDAASVNGIAKGIVDAEMKA----- 166
Db 1357 AKQALIDAA-----GDADAVDAKATAGIOA-----IDAKHKAGKTITDSRH 1395
Qy 167 -DAKEG-----KLNAAGAE-----TTNADAGKLFVKNAGV----- 197
Db 1396 DDAKQALIDEAAKVIKALIDODPTLTAQKEAKQAVATPADAKKALIDAGADAVDAK 1455
Qy 198 -----GEGAGDAK--AAAAVAASGEOLIAIVH-AAKDGEGEKGR--AAD 240
Db 1456 TAGIKALIDEOKHSGOTVDAKEDAKKALIDAEAG-KYTDALDHDTLTAQKEAKQAVAD 1514
Qy 241 RTNPIDAAIGAGDNDAA--AAAFATMKKDDQIAAAMVLRGMAKDDQ---FALKDAAAH 295
Db 1515 EADKAKKALIDAGNADALDQASAGIKALDE-----QHKSGQSIDTRKDDAKKALID 1565
Qy 296 GTVKNNAVDIKAAAEASASATGSAAGIDVYNGNGATATAGDAKGVNGIAK 348
Db 1566 GEVATITDAILDPTLTDAEKATQKQAVIAEADKAKKALIDAGDADAVDAK 1618

```

```

RESULT 7
US-09-500-746-17
; Sequence 17, Application US/09500746
; GENERAL INFORMATION:
; APPLICANT: Winchester, Robert J.
; APPLICANT: Gulko, Percio
; APPLICANT: Seki, Tetsunori
; TITLE OF INVENTION: USES OF INHIBITORS FOR THE ACTIVATION OF CXCR4 RECEPTOR BY SDF-1
; FILE REFERENCE: 0575/57005-B
; CURRENT APPLICATION NUMBER: US/09/500,746
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent version 3.1
; SEQ ID NO 17
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Human
US-09-500-746-17

```

Query Match 10.1%; Score 172.5; DB 5; Length 410;

Best Local Similarity 27.5%; Pred. No. 0.0036; Mismatches 161; Indels 97; Gaps 13;

```

Qy 3 NDHDHKGTVKNAVMAKAAFEASASATGNAAGIDVYNGGAAAKG--GEASVNGI 60
Db 58 NGNNNAGNANTGNTTGAAGANCAANGGGGNGTCGNNGAGCNGNTGTGAGTGG 117
Qy 61 ANGKIGYDAAG--KADAKEGKLDATGAE-----TTNNACKLFV 99
Db 118 AAGAAGGCNAAGCTCAANAAGAGCATATTTCAGANGNNGNAGGCGCTGTCNGC--- 173
Qy 100 KRAADGGADADAGKAAAVAAASAATGNAAGIDVYNGDVAKKAGGDAASVNGIAKGIC 159
Db 174 -----GGCAGTTTGTAAAAAAAANA--GACNG--CGACACACAAAGTGTNNCTGA 224
Qy 160 V---DAAEKADA--KEGKLNAAGAEETTNNADAGKLFVKNAGVNGEAGDAGKAAAVANS 215
Db 225 CCCGAAGCAANAGTGATNCAAGAGTACCTGG-----AGNNACTATGACAAANTAAC 278
Qy 216 GEOLIAIVHAKKDGEGEKKAADRTNPIDAAIG-----GAGNDAAAFAFTMKD 267
Db 279 GCAACAGCCAAAGAG-----ACTTNCCTAGACCCACGACGAGGAATAA----- 326
Qy 268 DOIIAAAMVLRGMAKDDQIAAAMVLRGMAKDDQ---FALKDAAAH 327
Db 327 -----AACCTTGT-----GAGAGATGAAGAGNCACAAAGAC 355

```

```

Qy 328 VNGNGATATAGK 338
Db 356 GTGGGGGAGCG 366

```

```

RESULT 8
US-09-249-585A-3
; Sequence 3, Application US/09249585A
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.0
; SEQ ID NO 3
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Epstein Barr Virus
US-09-249-585A-3

```

Query Match 10.0%; Score 171.5; DB 5; Length 641;

Best Local Similarity 27.2%; Pred. No. 0.0064; Mismatches 85; Conservative 3; Mismatches 151; Indels 73; Gaps 9;

```

Qy 27 SAASATGNAAGIDVYKNSGAAGGAAASVNGIAKGICIVDAAGKADAKEGKLDATGA 86
Db 89 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144
Qy 87 BETTVNAGKLFVKRAADGGADADAGKAAAVAAASAATGNAAGIDVYNGDVAKKAGGDA 146
Db 145 GCGAGAGCG--AGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 190
Qy 147 ASVNGIAKGICIVDAAEKADAKEGKLNAGAEETTNNADAGKLFVKNAGVNGEAGDAK 206
Db 191 GGAGGAGAG-----GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 229
Qy 207 AAAAANAASGEOLIAIVHAKKDGEGEKKAADRTNPIDAAIGGAGNDAAAFAFTMKK 266
Db 230 AGAGGAGAG-----AGAGGAGAG-----GAGGAGAGAGAGAGAGAGAGAGAGAGAG 264
Qy 267 DOIIAAAMVLRGMAKDDQIAAAMVLRGMAKDDQ---FALKDAAAH 326
Db 265 -----GAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 315
Qy 327 VVNGNGATATAGK 338
Db 316 GAGAGAGAGAGCG 327

```

```

RESULT 9
US-09-976-594-616
; Sequence 616, Application US/09976594
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 616
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

OTHER INFORMATION: Incyte ID No: 1692213CB1  
NAME/KEY: unsure  
LOCATION: 596  
OTHER INFORMATION: unknown or other  
US-09-976-594-616

Query Match 10.0%; Score 171; DB 5; Length 605;  
Best Local Similarity 30.2%; Pred. No. 0.0064;  
Matches 103; Conservative 7; Mismatches 163; Indels 68; Gaps 12;

OY 11 TKNVNDMAKAEFAASAATGNAAGDVYKNSGAAGGAASVNGIAGKIGYDA 70  
DB 164 TGTGAAAAACATGCGCAAAATCATGTGAA-----GCCATGAAAA-----CTTAA 206  
OY 71 AGKADAKEGKLDATGAECTTNNAGKLFVKRAADDGDAADGAKAAAVAAASAATGMAI 130  
DB 207 AGCTCG-----GATGAGATTCAATG-----AAGTGGCCGACAGTATAGTGAAGATAA 256  
OY 131 GDVNVGDVAKAKGDAASVNGIAGKIGYDAEKADEKGLNAAAG-----TTNA 184  
DB 237 GCCAGGCAAGGGGGTACTTG--GGTTGATGACCAAGAGGTCCATGTGGGACCATTTTC 314  
OY 185 DAGKLFVNAGNVGGEAGDAGKAAAVAAVSGEQLIKATVHAAKDGEKGGKAA--DR 241  
DB 315 AA-----GAGCAGCATTTGCCCTTCTGTA---AGTGGATGATTAAGCCTG 359  
OY 242 TNPIDAAIG-----AGNDAAAFATMKKDDQIAAAVLRGMARDDGFALDAAAHBG 296  
DB 360 TGTATACAGACCACCGGTAGA-----CAAAATTTGATATCATATTATTATG 410  
OY 297 TYKNAVDIKAAAEASAASATGSAATGIDVYNGKATAG 337  
DB 411 TCGAAGGAAGAAATTAATCATATGAAGACTG--AATAAG 450

RESULT 10  
US-09-952-267-7

Sequence 7, Application US/09952267  
GENERAL INFORMATION:  
APPLICANT: HANSEN, ERIC J.  
APPLICANT: AEBI, CHRISTOPH  
APPLICANT: COPE, LESLIE D.  
APPLICANT: MACIVER, MICHAEL J.  
APPLICANT: FISKE, MICHAEL J.  
APPLICANT: FREDENBURG, ROSS A.  
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS  
FILE REFERENCE: AMCY:024  
CURRENT APPLICATION NUMBER: US/09/952,267  
PRIORITY FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 09/336,447  
NUMBER OF SEQ ID NOS: 98  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 624  
TYPE: PRT  
ORGANISM: Moraxella catarrhalis  
US-09-952-267-7

Query Match 9.8%; Score 167.5; DB 5; Length 624;  
Best Local Similarity 25.4%; Pred. No. 0.01;  
Matches 89; Conservative 42; Mismatches 145; Indels 75; Gaps 17;

OY 9 KGTVNAV--DMAKAEFAASAATG--NAAIGDVYKNSGAAGGAASVNGIAGK 63  
DB 23 KHTTKAVIGSLIYCALGMATTASAGTGSTNAANGNIISVGAVYGGVINOAGKNPT 82  
OY 64 IKGIYD--AAGKADAKEGKLD--ATGAECTTNNAGKLFVKRAADDGDAADGAKAAAV 119  
DB 83 VCGGFDMNRATGNYSVYISGCFDNOAKGEHSTI-----AGGESNATGGRNSTV 128

OY 120 AASAATGNAAGDVYNGDVAKAKGDAASVNGIAGKIGYD-----AAEKADAKGKLN 174  
DB 129 A--GGSNNQAVG--TNSTYAGGSNNQAKANSFAAGVGNQANTDAAVAGKNTINGNNS 184  
OY 175 AA-GAEGTTNADAGKLFVNAGNVGGEAD-----AGKAAAV--AAVSG-----EQ 218  
DB 185 AAGSENTYVENOKNVFILLGSNTTNAOGSVYLLGHETSGKEATVAVSRAVNGITLKNFSG 244  
OY 219 ILKA--IYHAAKDGEKO-----GKKAADRTNPIDAAIGAGDNDAAAFATMKKDDQ 269  
DB 245 VKRADNGTVSVSGQEKEROIVHVAGQISDDSTD-----AVNGS---QLYALATAVDNQ 296  
OY 270 IAAAM-----VLRGMARDDGFALKDAAAAHGCTYKNAVDIKAAAE 310  
DB 297 YDIEINQDNIKDLQKEVKGDLKEVGLVSRDGLSLHDVDVADNADIAKKNKAD 347

RESULT 11  
US-09-952-267A-7

Sequence 7, Application US/09952267A  
GENERAL INFORMATION:  
APPLICANT: HANSEN, ERIC J.  
APPLICANT: AEBI, CHRISTOPH  
APPLICANT: COPE, LESLIE D.  
APPLICANT: MACIVER, MICHAEL J.  
APPLICANT: FISKE, MICHAEL J.  
APPLICANT: FREDENBURG, ROSS A.  
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS  
FILE REFERENCE: AMCY:024  
CURRENT APPLICATION NUMBER: US/09/952,267A  
PRIORITY FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: US/09/336,447A  
NUMBER OF SEQ ID NOS: 98  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 624  
TYPE: PRT  
ORGANISM: Moraxella catarrhalis  
US-09-952-267A-7

Query Match 9.8%; Score 167.5; DB 5; Length 624;  
Best Local Similarity 25.4%; Pred. No. 0.01;  
Matches 89; Conservative 42; Mismatches 145; Indels 75; Gaps 17;

OY 9 KGTVNAV--DMAKAEFAASAATG--NAAIGDVYKNSGAAGGAASVNGIAGK 63  
DB 23 KHTTKAVIGSLIYCALGMATTASAGTGSTNAANGNIISVGAVYGGVINOAGKNPT 82  
OY 64 IKGIYD--AAGKADAKEGKLD--ATGAECTTNNAGKLFVKRAADDGDAADGAKAAAV 119  
DB 83 VCGGFDMNRATGNYSVYISGCFDNOAKGEHSTI-----AGGESNATGGRNSTV 128  
OY 120 AASAATGNAAGDVYNGDVAKAKGDAASVNGIAGKIGYD-----AAEKADAKGKLN 174  
DB 129 A--GGSNNQAVG--TNSTYAGGSNNQAKANSFAAGVGNQANTDAAVAGKNTINGNNS 184  
OY 175 AA-GAEGTTNADAGKLFVNAGNVGGEAD-----AGKAAAV--AAVSG-----EQ 218  
DB 185 AAGSENTYVENOKNVFILLGSNTTNAOGSVYLLGHETSGKEATVAVSRAVNGITLKNFSG 244  
OY 219 ILKA--IYHAAKDGEKO-----GKKAADRTNPIDAAIGAGDNDAAAFATMKKDDQ 269  
DB 245 VKRADNGTVSVSGQEKEROIVHVAGQISDDSTD-----AVNGS---QLYALATAVDNQ 296  
OY 270 IAAAM-----VLRGMARDDGFALKDAAAAHGCTYKNAVDIKAAAE 310  
DB 297 YDIEINQDNIKDLQKEVKGDLKEVGLVSRDGLSLHDVDVADNADIAKKNKAD 347

RESULT 12  
PCT-US01-20545-4







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 24.93 Seconds

(without alignments)  
1066.383 Million cell updates/sec

Title: US-09-445-803-2

Perfect score: 1708

Sequence: 1 KNNDDHNNKCTVKNAMDMK.....GNCATAKGDAKSVNGIAGK 349

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.68:\*

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	266.5	15.6	353	140300	outer membrane lip
2	250	14.6	2639	T31328	fibroin - Chinese
3	225.5	13.2	555	S70791	vsb protein - Myc
4	224.5	13.1	362	I40304	outer membrane lip
5	219.5	12.9	246	M43579	Vmp7 protein homol
6	215.5	12.6	364	S11981	variable major pro
7	213.5	12.5	485	A60610	circumsporozoite p
8	208	12.2	1329	E70917	hypothetical glyci
9	205	12.0	1079	B70807	hypothetical glyci
10	204.5	12.0	398	OZ00AS	circumsporozoite p
11	204.5	12.0	1901	F70806	hypothetical glyci
12	202	11.8	1660	A70869	hypothetical glyci
13	201.5	11.8	429	A54504	circumsporozoite p
14	197	11.5	718	A36068	major amputate fi
15	197	11.5	1381	E70806	hypothetical glyci
16	197	11.5	1489	D70807	hypothetical glyci
17	195	11.4	1538	H70846	hypothetical glyci
18	194.5	11.4	973	C85693	probable membrane
19	193	11.3	837	E70835	hypothetical glyci
20	193	11.3	1011	F70620	hypothetical glyci
21	192	11.2	369	S11980	variable major pro
22	192	11.2	694	F70868	hypothetical glyci
23	191.5	11.2	798	T21369	hypothetical prote
24	191.5	11.2	1156	T34852	probable secreted
25	190.5	11.2	1122	G64887	probable tail fibe
26	186	10.9	882	B70812	hypothetical glyci
27	186	10.9	1207	T23754	hypothetical prote
28	185.5	10.9	606	H70816	hypothetical glyci
29	185.5	10.9	1306	A70934	hypothetical glyci

30	184.5	10.8	615	2	H70589	hypothetical glyci
31	183.5	10.7	801	2	F70824	hypothetical glyci
32	182	10.7	778	2	F70963	hypothetical glyci
33	182	10.7	2055	2	T31110	extracellular matr
34	181.5	10.6	421	2	JY0057	toia protein - Esc
35	181	10.6	532	2	F70580	hypothetical glyci
36	180.5	10.6	741	2	E70917	hypothetical glyci
37	179.5	10.5	767	2	E70895	hypothetical glyci
38	179	10.5	864	1	EART	elastin precursor
39	179	10.5	1822	2	S33441	EF protein - Stre
40	177	10.4	860	1	EAWS	elastin precursor
41	175.5	10.3	572	2	T08509	trid protein - Ent
42	175	10.2	498	2	C70720	hypothetical glyci
43	174.5	10.2	853	2	A70896	hypothetical glyci
44	173.5	10.2	591	2	B70523	hypothetical glyci
45	173.5	10.2	595	2	A48658	flagellin - Escher

#### ALIGNMENTS

```

RESULT 1
I40300
outer membrane lipoprotein - Borrelia hermsli
C:Species: Borrelia hermsli
C:Date: 12-Aug-1996 #sequence-revision 12-Aug-1996 #text-change 08-Oct-1999
C:Accession: I40300
R:Restrepo, B.I.; Kitten, T.; Carter, C.J.; Infante, D.; Barbour, A.G.
Mol. Microbiol. 6, 3299-3311, 1992
A:Title: Subtelomeric expression regions of Borrelia hermsli linear plasmids are high
A:Reference number: I40300; MUID:93133110
A:Accession: I40300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-353 <RES>
A:Cross-references: GB:I407488; NID:9144023; PIDN:AAA22963.1; PID:9144024
A:Gene: vmp17

Query Match          15.6%; Score 266.5; DB 2; Length 353;
Best Local Similarity 33.0%; Pred. No. 1.8e-07;
Matches 114; Conservative 36; Mismatches 122; Indels 73; Gaps 15;

QY 44 NSGAAGKGEAASVNGIANGIKGIIVD-----AAG-KDAKEGKLDA--TGAEGT 89
    ||| : : : || ||| : : ||| : : ||| : : ||| : : ||| : : |||
DB 20 NSGVAEDPKTYVLTJSIANLNGKGFLLVPTFGDMVVGARGIKADTKKSDIGKFTDIEST 79
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 90 -TNV-----NAGKLFVKRADDDGDADDAKGAANAASAAATGNAATGIVV--NGD 137
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 80 MTSVKKKLQDEVAKNNGNYPKVTAVDEFAVITGKIEKGAKEASKGATGVIIIGNTVKNGD 139
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 138 VAKAKGDAAASVNGIKGIYDAEKADEKADKGIINAAAGSETTNAADGKLFVKNAAGV 197
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 140 ---APEEATSVNSLVKGIKEIVGV---LKEGKADADATKDDSKDGIKLTATTTDAN 192
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 198 GGEAGDAGRAAANAASVSEGLKAIVHA-----AKDGEGKGRKAADRTNPIDAAIG 250
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 193 RADNMAAQAANAASIGAVTGADILQAIQSKENPVANSTGIEK-----ATDAALIAVA 245
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 251 GAGDN-----DAAAFATYKKDDQITAAAVLKGMAKDGAPALKDAAAAHGTVKNAVDII 305
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 246 PAKDNKEIKIDGA-----KKDAVIAAGIALRAMAANGTFISINBDA-----AVTTI 292
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 306 KAAAEASAAASATGSAATGIVV-----NGGATAKGDAKSV 343
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 293 NSRAASAVNKKIITLITAIKNTVDSGLKTIINEALATVKQED-KSV 336
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 2
T31328
fibroin - Chinese oak silkmoth
C:Species: Antherea pernyi (Chinese oak silkmoth)

```





QY 108 DADDAGKAAA VASAATGNA IGDVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKAD

Db 389 NQEEENDD\$NGPSEEHKKNYLESIRNSITEEWSPCSVTCGS 429

```

db      380 NOEENNDCNCRPEUTKVIESTBNCSTFEETWSDCVWCCS 420
      | |:: | |:: | |:: | |:: | |

```





C:Keywords: tandem repeat  
F:354-407/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 11.8%; Score 201.5; DB 2; Length 429;  
Best Local Similarity 30.2%; Pred. No. 0.00053;  
Matches 97; Conservative 26; Mismatches 135; Indels 63; Gaps 15;

QY 3 NDHDNHRKGVKNAVDMAKAAEEASASAAATGNAATGIVVKNKSGAAKGEASVNGIAK 62  
DB 114 NDAAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNA 172  
QY 63 GIKGIVDAAGKADAEGLKDATGAEETTVNNAKLFVKRAADDGDDADGKAAAVAAAS 122  
DB 173 G-----NAAGNAAGNAAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAAG 217  
QY 123 AATGNAATGIVVNGVYAKKAGGDAASVNGIATGIVDAEKADEKELNAAGAEETT 182  
DB 218 NAGNAA--GNAAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAG 265  
QY 183 NADAGKLFVKNAGNNGVGEAGDAGKAAA--VAAVSGEQLIKATVHAAKDGCKRKAADR 241  
DB 266 NA--AGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAG 320  
QY 242 TNPDAATGAGDNDAAAFAATMKDDQITAAAMVLRGMAKDCQFALKDAAAHEGTAKNA 301  
DB 321 DKNVDA-----NTNKKDNO-----EENNDSSNGPSEEH---IKNY 352  
QY 302 VDLIK-AAEAASASAAATGS 321  
DB 353 LESIRNSITEEMSPSCVTCGS 373

## RESULT 14

A36068  
major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)  
C:Species: Nephila clavipes  
C:Date: 08-Mar-1991 #sequence\_revision 13-Jan-1993 #text\_change 09-Sep-1997  
C:Accession: A36068  
R:Yu, M.; Lewis, R.V.  
Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990  
A:Title: Structure of a protein superfiber: spider dragline silk.  
A:Reference number: A36068; MUID:90384959  
A:Accession: A36068  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-718 <XU>  
A:References: GB:M37137; NID:9159711; PID:9159712  
A:Notes: the authors translated the codon GGT for residue 292 as Gln, GTA for residue 374

Query Match 11.5%; Score 197; DB 2; Length 718;  
Best Local Similarity 26.1%; Pred. No. 0.0014;  
Matches 98; Conservative 32; Mismatches 164; Indels 82; Gaps 13;

QY 20 KAAEENASASAAATGNAATGIVVKNKSGAAKGE-----ASVNGIATGIVDAAK 73  
DB 234 QGAGGAGAGSAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 293  
QY 74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGDDADGKAAAVAAASA----- 123  
DB 294 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 353  
QY 124 -----ATGNAATGIVVNGDVAKAKGDDA-----SYNGIATGIVDAEKADEK 170  
DB 354 GGLGSGAGRGGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 409  
QY 171 GRLNAGAEGTTNADAGKLFVNAGVGEAGDAGKAAAVAAVSGEQLIKATVHAAKDG 230  
DB 410 GGAAGRGYGTGLOAGAGR-----GGLGAG--AGAAAAAGAGAGGAGGAGGAGGAG 458  
QY 231 GEGGKKAADRTNPIDATGAGDN-----DAAAFAATMKDDQITAAA 273

DB 459 AGRGGGAA-----AAAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 510  
QY 274 MYLRGMAKDGQFALKDAAAHGTVKNAVDIITKAAAEASASAAATGSAATGIVVNGNA 333  
DB 511 ---QGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 557  
QY 334 TAKGDAKSVNGIAK 349  
DB 558 QGAGAGAAAAAGGVRGQ 573

## RESULT 15

E70806  
hypothetical glycine-rich protein RV3507 - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: E70806  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Felkell, T.; Gentles, S.; Hamilton, N.; Holtroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: E70806  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1381 <COI>  
A:Cross-references: GB:AL020222; GB:AL123456; NID:93261554; PIDN:CAA17744.1; PID:9292  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3507  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.5%; Score 197; DB 2; Length 1381;  
Best Local Similarity 29.1%; Pred. No. 0.0026;  
Matches 107; Conservative 14; Mismatches 163; Indels 84; Gaps 16;

QY 25 AASASAAATGNAATGIVVKN-----SGAAKGEBAASVNG-----IAKGITGIVDA 71  
DB 201 APCAAGCTGNGNGNALLIGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 260  
QY 72 GKADAEKGLDATTGAE--TNNVNAKLFVKRAADDGDDADGKAAAVAAASATGNAI 130  
DB 261 GTGGGCTAGCAAGAGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 318  
QY 131 GGVNGDVAKAKGDAASV---NGIATGIVVDAEKADEKELNAAGAGETTNADAG 187  
DB 319 GG--GGGDTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 369  
QY 188 KLFVNAGVGEAGDAGKAAAVAAVSGEQLIKATVHAAKGCKRKAADRTNPIDA 247  
DB 370 ---TGEGGTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 408  
QY 248 ATGAGDNDAAA-----FATMK--KDPDIAAMVLRGMAKDCQFALK----- 288  
DB 409 -TGCTGCACDAAYVFGAGNDPGFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 467  
QY 289 -----DAAAHEGTVKNAVDIITKAAAEASASAAATGSA--IGD-----VYNGCATAKG 337  
DB 468 GAGAGDAGSTGNPGKGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 527  
QY 338 GDAKSVNG 345  
DB 528 G-----VNG 531

Search completed: January 10, 2002, 14:04:28  
Job time: 122 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:04:31 ; Search time 15.37 Seconds

(without alignments)  
832.533 Million cell updates/sec

Title: US-09-445-803-2

Perfect score: 1708  
Sequence: 1 KNNDDHNRKGVAKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwisProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266.5	15.6	353	VM17_BORHE	P32777 borrelia he
2	224.5	13.1	362	VM25_BORHE	P32779 borrelia he
3	215.5	12.6	364	VM21_BORHE	P21875 borrelia he
4	213.5	12.5	393	CSP_PLABR	P15593 plasmidium
5	209.5	12.3	398	CSP_PLACC	P08673 plasmidium
6	204.5	12.0	1901	Y208_MYCTU	O53553 mycobacteri
7	201.5	11.8	429	CSP_PLAMA	P13815 plasmidium
8	197	11.5	747	SPD1_NEPCT	P19837 nephila cia
9	192	11.2	369	VM07_BORHE	P21876 borrelia he
10	190.5	11.2	1120	STFR_ECOLI	P76072 escherichia
11	189.5	10.7	801	Y747_MYCTU	O53810 mycobacteri
12	182	10.7	778	Y034_MYCTU	P71933 mycobacteri
13	181.5	10.6	421	TOIA_ECOLI	P19934 escherichia
14	179	10.5	864	ELS_RAT	O99372 rattus norv
15	177	10.4	860	ELS_MOUSE	P54320 mus musculu
16	175	10.2	498	Y118_MYCTU	O56615 mycobacteri
17	173.5	10.2	914	MA22_MYCTU	O06794 mycobacteri
18	171.5	10.0	488	CY12_BOVIN	O28092 bos taurus
19	171.5	10.0	641	EBN1_EBV	P03211 Epstein-Bar
20	170.5	10.0	774	STF_LAMB	P03764 bacterioph
21	169.5	9.9	419	CSP_PLACC	P08676 plasmidium
22	169.5	9.9	3591	FHAB_BORPE	P12255 bordelella
23	168	9.8	401	CSP_PLACC	P08674 plasmidium
24	166	9.7	957	Y278_MYCTU	P56877 mycobacteri
25	164	9.6	543	YD25_MYCTU	O10637 mycobacteri
26	162.5	9.5	503	YD25_MYCTU	O50630 mycobacteri
27	161.5	9.5	627	SPD2_NEPCT	P46804 nephila cia
28	161	9.4	542	PESL_ONCHY	P12027 onchophila
29	159	9.3	2038	FSH_DROME	P13709 drosophila
30	158.5	9.3	1654	OMPB_RICRI	O53047 r outer mem
31	157.5	9.2	5263	FBOH_BOMMO	P05190 bombyx mori
32	156.5	9.2	515	Y140_MYCTU	O50594 mycobacteri
33	155	9.1	331	MACS_HUMAN	P29966 homo sapien

34	154	9.0	1608	1	HLVA_SERMA	P15320 serratia ma
35	153.5	9.0	1150	1	APMU_PIG	P12021 sus scrofa
36	152.5	8.9	2249	1	OMPB_RICRI	P15921 rickettsia
37	150.5	8.8	280	1	MACS_CHICK	P16527 gallus gall
38	150	8.8	378	1	CSP_PLACC	P08675 plasmidium
39	150	8.8	747	1	ELS_BOVIN	P04985 bos taurus
40	149.5	8.8	463	1	YA68_MYCTU	O53416 mycobacteri
41	148.5	8.7	1655	1	OMPB_RICCN	O9K43 r outer mem
42	144.5	8.5	763	1	HTR2_HALNI	O9H81 halobacteri
43	144.5	8.5	764	1	HTR2_HALSA	P71410 halobacteri
44	144	8.4	1706	1	CYVA_BORPE	P15318 bordelella
45	143	8.4	3726	1	TRX_DROME	P20659 drosophila

## ALIGNMENTS

RESULT	ID	VM17_BORHE	STANDARD:	PRT:	353 AA.
AC	P32777				
DT	01-OCT-1993 (Rel. 27, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				
DT	01-FEB-1994 (Rel. 28, Last annotation update)				
DE	VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 17 PRECURSOR.				
GN	VM17.				
OS	Borrelia hermsli.				
OG	Plasmid.				
OC	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.				
OX	NCBI_TaxID=140;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SSP. HSI SEROTYPE 17;				
RA	MEDLINE=9313110; PubMed=1484486;				
RX	Restrepo B.I., Kiltien T., Carter C.J., Infante D., Barbour A.G.;				
RT	"Subtelomeric expression regions of Borrelia hermsli linear plasmids are highly polymorphic";				
RT	Mol. Microbiol. 6:3299-3311(1992).				
CC	-I- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING FROM ONE SURFACE EXPOSED VMP TO ANOTHER.				
CC	-I- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID ANCHOR.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: L04788; AAA22963.1; -				
DR	InterPro: IPR000680; Borrelia.lipo.				
DR	InterPro: IPR001800; Lipoprotein_6.				
DR	Pfam: PF00921; Lipoprotein_2; 1.				
DR	Pfam: PF01441; Lipoprotein_6; 1.				
DR	ProDom: PD001149; Lipoprotein_6; 1.				
DR	PROSITE: PS00013; PROKAR LIPOPROTEIN; 1.				
KW	Outer membrane; Lipoprotein; Signal; Plasmid.				
FT	SIGNAL	1	18		
FT	CHAIN	19	353		
FT	LIPID	19	19		
FT	SEQUENCE	353 AA;	36460 MW;	88FAB9B8C619AE91 CRC64;	
QY	Query Match	15.6%	Score 266.5;	DB 1;	Length 353;
QY	Best Local Similarity	33.0%;	Pred. No. 6.7e-07;		
QY	Matches 114;	Conservative 36;	Mismatches 122;	Indels 73;	Gaps 15;
QY	44 NSGAARKGEASVNGIAGKGIIVD-----AAG-KDAKEGKLD--TGAEGT 89				
QY	111 1: : : 11 11 1: : : 111 1: : : 1 1 1				
QY	20 NSGVAEDPKRYVLTLSIANLKGQFLDVFTFGDMYVGAGIKADPKKSDIGKFTIEST 79				

[illegible]

RESULT	2	
VM25_BORHE		
ID	VM25_BORHE	STANDARD;
		PRT;
		362 AA

DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 25 PRECURSOR  
UNCLD

0N VMP23;  
 0S *Borrelia hermslii*.  
 0C plasmid.  
 0C Bacteria; Spirochaetales; Spirochaetaceae; *Borrelia*  
 0X MORI\_TaxID=140;  
 0X

RN [1] —  
RP SEQUENCE FROM N.A.  
RC STRAIN-SSP, H51 SEROTYPE 25;  
RA MEDLINE=93133110; PubMed=1484486;  
RX Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.:  
RT "Subtelomeric expression regions of *Borrelia hermslii* linear plasmids  
RL are highly polymorphic."; Mol. Microbiol. 6:329-331(1992).

CC -1- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING  
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID  
CC ANCHOR.  
CC -1- SIMILARITY: STRONG, TO VMP7 AND VMP21.

-1- SIMILARITY: STRONG, TO VMP7 AND VMP21.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; L04787; AAA22965.1; -.  
DR InterPro; IPR000680; Borrelia\_lipo.  
DR

DR	PIASM.	PF00921.	LIPROTEIN_2.	1.
DR	PROSITE:	PS00013.	PROKAR.	LIPROTEIN; 1.
KW	OUTER membrane:	Liprotein:	Signal;	Plasmid.
FT	SIGNAL	1	26	PROBABLE.
FT	CHAIN	27	352	VARIABLE MAJOR OUTER MEMBRANE
FT	LIPID	27	27	LIPROTEIN 25.
FO	SEQUENCE	362 AA.	3646 MW.	N-ACYL DILYSERIDE (PROBABLE)
				2AC3E5248D5637BD CRC64

Query Match	13.1%;	Score 224.5;	DB 1;	Length 362;
Best Local Similarity	28.4%;	Pred. NO. 7.3e-05;		
Matches 99;	Conservative 37;	Mismatches 148;	Indels 65;	Gaps 14;

```

0Y 20 KAAEEAASAAATGNAIGDVYNSGAAAKGGGAASVNGIANGIKIYVAAKADKEG 79
Db 30 OAVENGKDGAAATGSLSEVLVEVKSAB-----NAFYEMALVPDTLGRATKDT 82
0Y 80 KLDATGAEGLTNVYAGKLFVYRAADD-----GQDADADGAKAAAAYVAAASATGNA----- 128
Db 83 KKNVVG--GYNSLSLGLT--GKASDELEEVAKKVEYVGADGPIVAAIRAAVDTAKTLLS 138
0Y 129 -----AIGDVVN-GDVYAKKGGDAAS--VNGIANGIKIYVAAEKADKEGKL- 173
Db 139 TLKEHLESLKIGIDDDKRVGEATSNQNGVAASTDLTKAPFALKGIYDPAKGEVAKPKAG 198
0Y 174 --NAAGAEETTAAODGKLFVKNAGVGESEAGDAGCAAAAAYVAAVSEGLIKIYVAAKDGE 222
Db 199 DTAFKIGNADNKKDCAKYLAA--AANAGRAVGD--KAAIVASVASEEMLASTV--ASQEGD 253
0Y 233 KQCKAAADRTPIDALIGAGDNDNAAPFTMKRDDGIIAAMVLRGMAKRGCF----- 285
Db 254 ADAALADATPATQTSALKFARGGGAGAGOLAKEAAMAAAVAGGIALRSTLYVGKGLAANND 313
0Y 286 -----ALKDAAAANHEG-----TYKNVNDIIKAAEEAASAAASATG 320
Db 314 DKVYOSAGVTAVNKLVAAYBEIIGKTIYKNVLEKKKGGEIDAKARAK-ATG 361

```

Db 314 DKVQSGVTAVNKLVAVEGI IKTKNVLEKAKGEIDKARAPK-ATG 361

RESULT	3	
VM21_BORHE		
ID	VM21_BORHE	STANDARD;
		PRT;
		364 AA

AC P21875;  
DT 01-MAY-1991 (rel. 18, Created)  
DT 01-MAY-1991 (rel. 18, last sequence update)  
DT 01-FEB-1994 (rel. 28, last annotation update)  
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 21 PRECURSOR

GM  
OS  
OG  
OC  
OX  
NCHI\_TaxID=140;  
VMP21:  
Borrelia hermsli.  
Plasmid.  
Bacteria: Spirochaetales; Spirochaetaceae; Borrelia  
NCHI\_TaxID=140;

RN  
RP  
RC  
RX  
RA  
RT  
RQ  
RL

(1)

SEQUENCE FROM N.A.  
STRAIN-SSP. HSI SEROTYPE 21;  
MEDLINE-91171872; Pubmed-1706456;  
Burman N., Bergstrom S., Restrepo B.T., Barbour A.G.;  
"The variable antigens Vmp7 and Vmp21 of the relapsing fever  
bacterium Borrelia hermsli are structurally analogous to the VSG  
proteins of the African trypanosome.";  
Mol. Microbiol. 4:1715-1726(1990).

RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-SSP. HSI SEROTYPE 21;  
RA MEDLINE-93133110; PubMed-1484486;  
RX Restrepo R.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;  
RT "Subtelomeric expression regions of *Borrelia hermsli* linear plasmids  
are highly polymorphic.",  
RL Mol. Microbiol. 6:3299-3311(1992).  
NP  
(3)  
SEQUENCE OF 55-75; 186-208 AND 245-259.

RX  
MEDLINE=85236116; PubMed=24091997;  
RABEST P.A., Colligan J.E., Raam M.G., Barbour A.G.;  
RT "variable major proteins of *Borrelia hensili*. Epitope mapping and  
partial sequence analysis of CNBR peptides."; *J. Exp. Med.* 161:1302-1314(1985).

CC -1- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING  
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID  
CC ANCHOR.  
CC -1- SIMILARITY: STRONG, TO VMP7 AND VMP25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

RESULT	4
CSP_PLABR	
ID	CSP_PLABR
AC	PI4593;
DT	20-JAN-1990 (rel. 13, Created)
DT	20-JAN-1990 (rel. 13, Last sequence update)
DT	20-AUG-2001 (rel. 40, Last annotation update)
DE	CIRCUMSPOROZOITE PROTEIN (CS) (FRAGMENT).
OS	Plasmodium brasilianum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5824;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=88186854; PubMed=3128542;
RA	Lal A.A., Ia Cruz V.E., Collins W.E., Campbell G.H., Procell P.M.,
RA	McCutcheon T.F.;
RT	"Circumsporozoite protein gene from Plasmodium brasilianum. Animal
RT	reservoirs for human malaria parasites?";
RL	J. Biol. Chem. 265:5495-5498(1988).
CC	-1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC	SURFACE ANTIGEN ON THE SPOOROITE (THE INFECTIVE STAGE OF THE
CC	MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC	VERTEBRATE HOST).
CC	-1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC	ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC	WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC	-1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.

RESULT	5
ID	CSP_PLACC
AC	P08673.
DT	01-JAN-1988 (Rel. 06, Created)
DT	01-JAN-1988 (Rel. 06, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	PLASMODIUM POROZOITE PROTEIN PRECURSOR (CS).
OS	Plasmodium cynomolgi (strain Ceylon).
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5629;
RN	(1)
RP	SEQUENCE FROM N.A.
RX	MEDLINE=87102878; PubMed=3802196;
RA	Galsinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA	Nussenzweig R.S., Enea V.;
RT	"The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL	Cell 48:311-319(1987).
CC	-I- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC	SURFACE ANTIGEN ON THE SPOROZYTE (THE INFECTIVE STAGE OF THE
CC	MALARIA PARASITE) THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC	VERTEBRATE HOST).
CC	-I- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR

```

CC      ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC      WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC      -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC      -----
CC      CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      CC      the European Bioinformatics Institute. There are no restrictions on its
CC      CC      use by non-profit institutions as long as its content is in no way
CC      CC      modified and this statement is not removed. Usage by and for commercial
CC      CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: M15103; AAA29533.1; -.
DR      PIR: C6255; OZDOAS.
DR      InterPro: IPR003067; Circmsprzoite.
DR      InterPro: IPR000884; TSP1.
DR      Pfam: PF00090; tsp_1; 1.
DR      PRINTS: PR01303; CIRCMSPRZOITE.
DR      SMART: SMO0209; TSP1; 1.
DR      K12 Malaria; Sporozoite; Repeat; Signal.
FT      SIGNAL 1 19
FT      CHAIN 20 398
FT      DOMAIN 97 249
FT      CIRCUMSPOROZOITE PROTEIN.
FT      17 X 9 AA TANDEM REPEATS OF A-G-N-A-A-
FT      A-G-E.
SEQUENCE 398 AA; 37718 MW; 6DFAZE8A62ED05BF CMC64;

```

Query Match	12.3%;	Score 209.5;	DB 1;	Length 398;
Best Local Similarity	27.2%;	Pred. NO. 0.00042;		
Matches 97; Conservative	21;	Mismatches 129;	Indels 109;	Gaps 11;

```

QY 7 NHKTVKNAVYMAAAAEBAASAAATGNAALIGVYVKKSGAAAGGEAAVSYNGIAKIGK 66
Db 33 NLNGVFNINVASSLG--AAOVROSASRKRGLGENPPKNEGA-----72

QY 67 IVDAAKADAK-----EGKL-----DATGEGTTNNAGKLFYRAADCGDADA 1127
Db 73 --DKPKKDEKQVPEKKRREKTKLQPNAGNNAAAEAGNNAAAGGAGNNAAA 1304

QY 113 GKAAAAVAASAATGNAALIGDVNGDVAKAGGDA--SYNGIAKIGIYDAAEKAD 165
Db 131 GEAGNNAAAGGAGNNAAAGGAGNNAAAGCAAGNNAAAGGAGNNAAAGGAAG-----NNAAG 188

QY 170 EGTKLNAGAEBOITTAADAGKLEVYKAGVNGGE--AGDAKRAAAVAAVSEQIILKATVHA 222
Db 186 EAGNNAAGGEGAGNNAAAGGAGNNAAAGGAAGNNAAAGGAGNNAAAGGAAGN-----234

QY 227 AKDGGEGKQKKAADRTNPIDAAITGAGDGNDAAAAFATMKKDDQITAAAMVLRGMAKDGQFA 287
Db 235 -----NNAAGAGAGNNAAAG-----248

QY 287 LKDAAAAEHTVKNNAVDILIKAAAEASA--SAATGSAALGDVYVNGCATKATKAGDA 340
Db 249 --EAGAGAGAGNNAAGGAGAGAGACAGACNNAAAGGAGAGC--AGNAGNKKRAGDA 301

```

RESULT	6
YZ08_MYCTU	
ID	YZ08_MYCTU STANDARD: PRT, 1901 AA.
AC	053553;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	20-APR-2001 (Rel. 40, Last annotation update)
DE	HYPOHETICAL PE-PCRS FAMILY PROTEIN RV3508 PRECURSOR.
OS	RV3508 OR MT023.15.
GN	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium
OX	NCBI_Taxid=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLaren J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Ratandream M.A., Rogers J.,  
RA Sulten S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,  
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
PL complete genome sequence." *Nature* 393:537-544(1998).

```
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL0202022; CAA17745.1; -.
DR HSSP; P19972; IKVD.
DR TubercuList; RV3508; -.
DR InterPro: IPR0000084; PE.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE; 1.
DR Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHEITICAL PE-PGRS FAMILY PROTEIN
FT RV3508
SQ SEQUENCE 1901 AA; 147627 MW; C7BI923D5D0146CD CRC64;
```

Query Match	12.0%;	Score 204.5;	DB 1;	Length 1901;
Best Local Similarity	27.6%;	Pred. NO. 0.0029;		
Matches 106; Conservative	25;	Mismatches 136;	Indels 117;	Gaps 19

QY	10	GTVAKNA-----	VDYMKAAEEAASAAATGNAIGDIVYKNSGNAAKGFEASVNG	59	
Db	668	GSVGNAGIGGTGGTGGYGGAGACAAAAGSSATGCGFA-----	GCAGEGC-AGGNSG	7231	
QY	60	IANKIGIVDAACAKADAKEGTLDTATGEGTTNNVAMGKLFYKRAADDCGDADDGKAAAAV	119		
Db	722	V-----	GGTNGSGGAGGACGCGTGGAGGSGADN-----PTGAGFAFGAGGTGGAAAG	770	
QY	120	AASATGNAATGDIVY-----	NGDVAKAKGGAASVNGIAKTIKGIVDAAE-	164	
Db	771	GAGGATGGTGGYVVGATGSAGTGGAGRGD-----	GCGGAS--GIGLGLISGF-DGCGG	8322	
QY	165	-----	KADAREGKLTNAAGAEETTNAADG-----	KLIFYKMNNGVSGEFGDAGSKAA	208
Db	823	GCGGAGGSAGAGGGINGAGAGCGCGDGDGATGAAGLGDNGYGGCGDGCAGGAAGCGNAG	8622		
QY	209	AAVAASGEQILKAIVHAAKDGCCEGCKKAADPTNPIDAAIGCAGDIDAATAAFAFMKKDD	268		
Db	883	VGLTAAKXG-----	DGCG-----AAGNGGNISGAGGAGGAGDNNFMNG-----	917	
QY	269	QIAAAMWLRMA--KDCQFALKDAAA-----	AHECTYKNAVDIIKAAAEASAAASAATGSA	3322	
Db	918	-----	QGGAGCGGCGGCGGCTISINANGAGCGNG-----GTGGKGGAGGAG	960	
QY	323	ALGPDVNGNCAT-AKGGDAKSVNG	345		
Db	961	TLG--VGSGGCTGGDGDAGSGCG	982		

RESULT	7
CSP_PLAMA	
ID	CSP_PLAMA
AC	P13815;
DT	01-JAN-1990 (rel. 13, Created)
	STANDARD;
	PRT; 429 AA



```

DT 01-JAN-1990 (Rel. 13, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS Plasmodium malariae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5858;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040027; PubMed=3054537;
RA Lal A.A., Ja Cruz V.F., Campbell G.H., Procell P.M., Collins W.E.,
RA McCutchan T.F.;
RT "Structure of the circumsporozoite gene of Plasmodium malariae.";
RT Mol. Biochem. Parasitol. 30:291-294(1988).
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J03992; AAA29557.1; -
DR PIR: A54504; A54504.
DR HSP: P04002; IWEA.
DR InterPro: IPR003067; Circmsprzoite.
DR Pfam: PFO0090; tsp_1; 1.
DR PRINTS: PR01303; CIRCMSPRZOITE.
DR SMART: SM00209; TSP1; 1.
DR PROSITE: PS50092; TSP1; 1.
DR Malaria: Sporozoite; Repeat; Signal.
FT SIGNAL 1 15 PROBABLE.
FT CHAIN 16 429 CIRCUMSPOROZOITE PROTEIN.
FT DOMAIN 110 314 4 AA TANDEM REPEATS.
SQ SEQUENCE 429 AA; 41596 MW; 3629D641D1C0BB7E CRC64;

```

```

RESULT 8
ID SPDL_NEPCL STANDARD; PRT; 747 AA.
AC P19837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, last sequence update)
DT 01-OCT-1996 (Rel. 34, last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
RT proteins (spidroin) from Nephila clavipes (Tetragnathidae) and
RT Araneus bicentenarius (Araneidae).";
RT J. Biol. Chem. 269:6661-6663(1994).
CC -1- FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE
CC CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
CC PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
CC WITH ELASTIC AMORPHOUS SEGMENTS.
CC -1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M37137; AAA29380.1; -
DR EMBL: U03848; AAB60212.1; -
DR PIR: A36068; A36068.
KW Silk; Repeat.
FT NON-TER 1 1
FT DOMAIN 1 655 25 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 1 1 25
FT REPEAT 26 38 2.
FT REPEAT 39 66 3.
FT REPEAT 67 96 4.
FT REPEAT 97 130 5.
FT REPEAT 131 158 6.
FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.

```



OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aliba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Salto N.,  
RA Sempel G., Seki Y., Sivasubaram S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AE000234; AAC74454.1; ALT\_INIT.  
DR EMBL: D90774; BAA14966.1; -  
DR EMBL: D90775; BAA14975.1; -  
DR HSP: F04002; IMFA.  
DR Ecocore: EG13370; stfr.  
DR InterPro: IPR000122; Chemotaxis\_transducer.  
DR Hypothetical protein; Fiber protein; Repeat; Complete proteome.  
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

Query Match 11.2%; Score 190.5; DB 1; Length 1120;  
Best Local Similarity 24.9%; Pred. No. 0.0085;  
Matches 92; Conservative 47; Mismatches 168; Indels 63; Gaps 12;

Q 6 DHKGTAVNAVDMKAAEFAASAASANTGMAIGDVKNKSGAAKCGEASVNGIAKGIK 65  
D 191 ESSKSAATSAQAAGTSETNASASQSAATSTATTKASEATSRDAASKEAKSSE 250  
Q 66 GIVDAAGKADAEKGDANG-----AECTNVNAGKLFVRRADDGDADAGAAVA 121  
D 251 --TNASSSSAASSATAATGNSAKAKTSETNARS--SETAAGQSAASAAGSTAAASSA 306  
Q 122 SAATGNAAGIDVNGDVAKAGGDAASVNGIAKGIKGVDAEAKADAKGKINAAGACT 181  
D 307 SAASTSA-----GQASASATAAGKSAESAASAASASTATTKGEATEQASAA 352  
Q 182 TNADAGKLFVNAGNVGEGAGDAARAAVAAVSEGOILKATVHAAKDGEGOKKRA-AD 240  
D 353 RSASAKTSETNA-KASETSAESSKTAASAASASSASS--ASASKDEATROASAKSS 410  
Q 241 RTNPIDAAIGAGGDDAA-----AFATPKKDDOFAAANVLRGMAKDDGPAIK 288  
D 411 ATTAATKATEAAGSATAAQSKTASATRAETAKRAEDTASAV-----ALE 459  
Q 289 DAAAAHEGTAVNAVDIKAARAAASAASAAATGSAAGIDVNG-----NGATA--KGG 338  
D 460 DASTTKKGIIVQ-----LSSATNSTSETTLATPKAVKSAVDNMEKRLQKQNGADIPDKGC 514

QY 339 DAKSVNGIAK 348  
DB 515 PLNNINAVSK 524

RESULT 11  
Y747\_MYCTU STANDARD: PRT; 801 AA.  
ID Y747\_MYCTU  
AC 053810;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PE-PCRS FAMILY PROTEIN RV0747 PRECURSOR.  
GN RV0747 OR MT0772.5 OR MTV041.21.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean L., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,  
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CDC 1551 / oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey S.L.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PCRS  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AL021958; CAI1514.1; -  
DR EMBL: AE006968; AAK45011.1; -  
DR TIGR: MT0772.5; -  
DR Tuberculist: RV0747; -  
DR InterPro: IPR000084; PE.  
DR Pfam: PF00934; PE; 1.  
KW Hypothetical protein; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 30  
FT CHAIN 31 801  
FT FT  
FT CONFLICT 188 188 G -> S (IN REF. 2).  
FT CONFLICT 225 225 R -> G (IN REF. 2).  
FT CONFLICT 227 227 R -> G (IN REF. 2).  
FT CONFLICT 295 295 K -> R (IN REF. 2).  
FT CONFLICT 300 300 S -> G (IN REF. 2).  
FT CONFLICT 338 338 T -> I (IN REF. 2).  
FT CONFLICT 377 377 A -> P (IN REF. 2).  
FT CONFLICT

FT CONFLICT 577 577 T -> A (IN REF. 2).  
 FT CONFLICT 580 595 MISSING (IN REF. 2).  
 SO SEQUENCE 801 AA; 65407 MW; EA54C9BFA5A00F41 CRC64;

Query Match 10.7%; Score 183.5; DB 1; Length 801;

Best Local Similarity 27.0%; Pred. No. 0.014;  
 Matches 107; Conservative 29; Mismatches 162; Indels 99; Gaps 19;

19 AKAEEAASAS-----AATGNAIGD-----VKNNSGAAG 51  
 86 AAAAAAVALPILPINAQFAATGRPLTGANGAPCTGANGPGMLJGNGAGSG 145  
 52 --GEAASVNGIAKIGIYDAAG--KADAKEGLDATTGAGTTNVAGKLFVRADDG-- 106  
 146 AAGAGAGNGAGAGLFGSGGAGASTDVAGAGAGAGAGG-----NAGMLF--GAAGVGCV 199  
 107 -----GDDDDAKAAAATAAATGNAIGDVNGDVAKKAGDA--ASVNGIAK 154  
 200 GGFSGNGATGGAGAGAGLFGAGRENGSGSGNLTG--AGGAGNAGTATGPDGAG 257  
 155 GIKGIYDAEAKADAKGKLNAAAGETTNADAGKLFVK-----NAGNVGEAGD--AGKAA 208  
 258 GTGGA-----SRSGGGAGAG--GDAGMFFGSGSGAGGTSKSVGDSAGAGAG 306  
 209 AAVAAV-----SGEQLIKAIYHAADGEGEKGRKAADRTNPIDAIAGAGDNDPAAAF 261  
 307 GAGPLGNGGNGNG-----GASTGGDGGPGGAGCTVL--TGNGNGSGCGTC 354  
 262 ATMKDDQIAAAMVLRGMAKDGQFALKDAAAHE-----GTVKNVNDIKKAAEAASAA 315  
 355 ATLGKAGIGCTGGVLLGL--DEFTAPASTSPHTLQDDVINYNDPFTLTGRPLGNCA 412  
 316 SAATGSAAGDV--VNGNGATAGKGDASKVNGIAK 349  
 413 NGPTGTAGDAGAGCMLEFGNGNGCGCTIGVNGAGAG 449

RESULT 12  
 YQ34\_MYCTU STANDARD; PRT; 778 AA.  
 ID YQ34\_MYCTU  
 AC P71933;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.  
 GN RV2634C OR MT2712 OR MTCY441.04C.  
 CC Mycobacterium tuberculosis.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RC MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares S.,  
 RA Sultun J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
 CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: Z80225; CAB02341.1;  
 DR EMBL: AE007103; AA047026.1; ALT\_INIT.  
 DR HSSP: P04002; IATP.  
 DR TIGR: MT2712;  
 DR Tuberculist: RV2634C;  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR000084; PE.  
 DR Pfam: PF00934; PE.1.  
 DR Prodom: PD001223; PE.1.  
 DR Hypothetical protein; Complete proteome.  
 KW CONFLICT 51 V -> L (IN REF. 2).  
 FT CONFLICT 51 V -> L (IN REF. 2).  
 FT CONFLICT 63 Q -> H (IN REF. 2).  
 FT CONFLICT 274 A -> T (IN REF. 2).  
 SO SEQUENCE 778 AA; 63131 MW; DAB20FE5B64999E7 CRC64;

Query Match 10.7%; Score 182; DB 1; Length 778;  
 Best Local Similarity 26.6%; Pred. No. 0.016;  
 Matches 89; Conservative 24; Mismatches 130; Indels 92; Gaps 17;

29 ASAATGNAATG--VKNNSGAAGGGAASVNGIAKIGIYDAGKAD---AKEGLDA 83  
 314 AGAAGNGAGAGMLLGDGAGGCGGPA--VAGVLGMPAGGNGNANMFGSGGCGGCG 371  
 84 TGAEGTTNVNACKLFLVRRADDDGADAKAAVAASATGNAATGIVNGDVAKKAG 143  
 372 TGLAGTNGVNPESIANPNTGANGTD-----NSGNGNOTG--NGGEPAGG 415  
 144 -GDAASVNGIAKIGIYDAEAKADAKGKLNAAAGETTNADAGLFLVKNAGNVGGEAG 202  
 416 VEGAGVGCG--GGGCG-----ESLDONDTGGKGGAGCTAGTDGAGAGAGGAGIGETDG 468  
 203 DAGKAAAANAASGEQILKAIYHAADGEGEKGRKAADRTNPIDAIAGAG-----DN 255  
 469 SAGGVAT-----GGE--GDXA--TGGVDDGVGAGGKGGCGHNT 504  
 256 DAAAFATMKDDQIAAAMVLRGMAKDGQFALKDAAAHEGTIVKNVNDIKKAAEAASAA 315  
 505 GVGDAFG-----GD-----GIGGDCNGAL--GAAGNGGT-----GAGG 538

316 SAATGSAAGD-----VNGNGATAGKGDASKVNGI 346  
 539 NGRGRGMLJGNGAGAGAGTGTGGGAGGAFGAGV 573

RESULT 13  
 TOLA\_ECOLI STANDARD; PRT; 421 AA.  
 ID TOLA\_ECOLI  
 AC P19934;  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TOLA PROTEIN.  
 GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-JM105:  
RX MEDLINE-90078104; PubMed-2687247;  
RA Levengood S.K., Webster R.E.;  
RT "Nucleotide sequences of the *tolA* and *tolB* genes and localization of  
their products, components of a multistep translocation system in  
Escherichia coli.";  
RL J. Bacteriol. 171:6600-6609(1989).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN-K12;  
RX MEDLINE-97061202; PubMed-8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
[4]  
RN DOMAIN.  
RP MEDLINE-91296736; PubMed-2068069;  
RX Levengood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "TolA: a membrane protein involved in colicin uptake contains an  
extended helical region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
[5]  
RN INTERACTION WITH PORINS.  
RP MEDLINE-97133271; PubMed-8978668;  
RX Derouiche R., Gavioi M., Benedetti H., Prilipov A., Lazdunski C.,  
Lioudes R.;  
RT "TolA central domain interacts with Escherichia coli porins.";  
RL EMBO J. 15:6408-6415(1996).  
[6]  
RN X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.  
RP MEDLINE-99332679; PubMed-10404600;  
RX Subkowski J., Hennecke F., Plueckthun A., Wlodawer A.;  
RT "Filamentous phage infection: crystal structure of g3p in complex  
with its coreceptor, the C-terminal domain of TolA.";  
RL Structure 7:711-722(1999).  
CC -I- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A  
COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE  
COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL  
BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION  
OF BACTERIOPHAGE DNA.  
CC -I- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPG, PHOE  
AND LAMB.  
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: M28232; AAA24683.1; -  
CC EMBL: AE000177; AAC73833.1; -  
CC EMBL: D90713; BAA35405.1; -  
DR

DR PIR: JY0057; JY0057.  
DR PDB: 1TOL: 20-MAY-99.  
KW Ecogene: Ec11007; tolA.  
KW Transport: Protein transport; Bacteriocin transport; Transmembrane;  
KW Repeat: Inner membrane; 3D-structure; Complete proteome.  
FT DOMAIN 1 13  
FT TRANSMEM 14 34  
FT DOMAIN 35 421  
FT DOMAIN 35 421  
FT DOMAIN 48 310  
FT DOMAIN 311 421  
FT DOMAIN 224 278  
FT DOMAIN 224 278  
SO SEQUENCE 421 AA; 43156 MW; 882F52B4B97C655E CRC64;  
Query Match 10.6%; Score 181.5; DB 1; Length 421;  
Best Local Similarity 27.4%; Pred. No. 0.0099;  
Matches 101; Conservative 44; Mismatches 141; Indels 83; Gaps 15;  
QY 18 MAKAEFAASASATGNAIGDV-----KNSGAACKGGAASVNGIAKIRGI 67  
DB 1 MSKATEQNDKLRRAIISVHLVILFALIWSEFDENIEASGGGGSSIDAVMVDGAV 60  
QY 68 VD-----AAGKADKEGKLDATGAGETTNVAG-----KLFVR--AADGCGDA 109  
DB 61 VEQYKRMQSOESSAKRSQKKEQQAABELREKQAAQERLKQLEKERLANQEQKQA 120  
QY 110 DDAGKAA-----AAVAAATGNAIGDVNGVAKAG-GDAASVNGIAKIRGIDA 162  
DB 121 EEAAQAEELKQQAEEAAKAAAD-----AKAAEADAKAAEEAK--KAAADA 167  
QY 163 AEKADAEKELNAGAEETTNADAGKLFYKMGNGVGEAGDAGKAAAVAVSGEOLKA 222  
DB 168 KKAEEAEAAK-----AAEAQKAAEAAAALK-----KKEEAEAAAEARKKA 211  
QY 223 IYHAAKDGEGKQKRAADTNPIDAIGAGDNDAAAFATM-KKDDQIAAAMVLMGMAK 281  
DB 212 ATEAAEKAAAEKKA-----AAEKNAADKAAAEKAAADKKAEEKAA---EKAAD 260  
QY 282 DQGFALKDAAAEHETGVKNAVDIIKAAAEASASATGSAI-GDVNGCGATKAGGDA 340  
DB 261 DKKAAEKAAADKKA-----AKAAAEKAAAKAAAEADDFLGELSSGKNPKTGGCA 314  
QY 341 KSVNGIAKG 349  
DB 315 KGNNAAPAG 323  
RESULT 14  
ELS\_RAT  
ID ELS\_RAT STANDARD: PRT: 864 AA.  
AC 099372;  
DT 01-OCT-1996 (rel. 34, Created)  
DT 01-OCT-1996 (rel. 34, Last sequence update)  
DT 20-AUG-2001 (rel. 40, Last annotation update)  
DE ELASTIN PRECURSOR (TROPOLASTIN) (FRAGMENT).  
GN ELN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91104868; PubMed-1702999;  
RA Pierce R.A., Deak S.B., Stollie C.A., Boyd C.D.;  
RT "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";  
RL Biochemistry 29:9677-9683(1990).  
[2]  
RN SEQUENCE OF 781-864 FROM N.A.  
RP MEDLINE-88330868; PubMed-2971041;  
RX Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;  
RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";  
RL J. Biol. Chem. 263:13504-13507(1988).  
DR

[3]  
 RN SEQUENCE OF 264-533 AND 558-864 FROM N.A.  
 RX MEDLINE=92241859; PubMed=1572637;  
 RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;  
 RT "Elements of the rat tropoelastin gene associated with alternative  
 RT splicing.";  
 RL Genomics 12:651-658(1992).  
 CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND  
 CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
 CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
 CC INTO AN EXTENSIBLE 3D NETWORK.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.  
 CC -1- ALTERNATIVE PRODUCTS: THREE DIFFERENT MRNAs HAVE BEEN FOUND THAT  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M60647; AAA42269.1; -  
 DR EMBL; J04035; AAA42268.1; -  
 DR EMBL; M86372; AAA42271.1; -  
 DR EMBL; M86355; AAA42271.1; JOINED.  
 DR EMBL; M86363; AAA42271.1; JOINED.  
 DR EMBL; M86364; AAA42271.1; JOINED.  
 DR EMBL; M86366; AAA42271.1; JOINED.  
 DR EMBL; M86371; AAA42271.1; JOINED.  
 DR EMBL; M86376; AAA42272.1; -  
 DR EMBL; M86373; AAA42272.1; JOINED.  
 DR EMBL; M86375; AAA42272.1; JOINED.  
 DR HSSP; P04002; IWPA.  
 DR Structural protein; Connective tissue; Repeat; Signal;  
 KW Alternative splicing.  
 FT NON\_TER 1  
 FT SIGNAL 1  
 FT CHAIN 22 864 BY SIMILARITY.  
 FT VARIANT 263 307 MISSING (IN CERTAIN CLONES).  
 FT VARIANT 308 308 MISSING (IN CERTAIN CLONES).  
 FT VARIANT 809 823 MISSING (IN CERTAIN CLONES).  
 FT SEQUENCE 864 AA: 72786 MW: 456894BB09E79FD4 CRC64;  
 SO

Query Match 10.5%; Score 179; DB 1; Length 864;  
 Best Local Similarity 26.3%; Pred. No. 0.024;  
 Matches 113; Conservative 20; Mismatches 192; Indels 104; Gaps 14;

19 AKAEEAASAA--SAATGNAAGDVVKNKSGAAAGK-----GEAASVN 58  
 346 AAAAAAATAAATGAAAGLVPGGPRVPGAGIPGVGIGIPVGGIPVGGIPG 405  
 59 GIATKIGKIVD-----AAGKADAKKEGLDATTGAE--TTNVNAGKL--FVKRAAD 105  
 406 GPGIGPGIPGPGAVSPAAAAAKAAKAYGARGVGIPTYGVGAGCPGPGYGAG 465  
 106 GGDADAGCAAAVAASATGNA-AIGDVVNDVAKA-----KGDAAASYNGIA 155  
 466 GGAASAAAAAATAAATGAAAGLVPGGPRVPGAGIPGVGIGIPVGGIPG 525  
 156 IKG-----IYDAEKADAKKEGL-----NAAAGAGTTN 183  
 526 VPGTGVPGAGTPAAAAAATAAATAAAGAGVGLGPGVGVGGLGPGVGGVGTG 585  
 184 ADAGKLFYKNAGNVGEGADGAKAAAVAAVAGSEQLKAIYHAARD-GEKOGKRAADPT 242  
 586 IGTG-----PGTGLVPGDAGGAGTTPAAAKSA--AKAAAKAYRAAAGLGGVGLGVAGV 639  
 243 NPIDAIGAGAGNDAAAFATYTKKDDIATAANVLR-----GMAKKDQGF 285

DB 640 PGFGAGGCGFGAGVPGFGAGAVPGSLAASAKAYGAGGLGPGGLGPGG 699  
 QY 286 ALKDAAAHEGVYKNKAVNDIIRKAAEFAASAS-----AATGSAAGVYVNGNATAGGSA 340  
 DB 700 GGPBGGLGVPGVPGVAGAPAAAAAATAAATAAAYGLGAGGLGAGGLGAGG 759  
 QY 341 KSYNGIANG 349  
 DB 760 LGAGGLGAG 768

RESULT 15  
 ELS\_MOUSE  
 ID ELS\_MOUSE STANDARD: PRT: 860 AA.  
 AC P54320;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE ELASTIN PRECURSOR (TROPOELASTIN).  
 GN ELN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE:Lung;  
 RX MEDLINE=95130069; PubMed=7829060;  
 RA Wyther K.S., Sechler J.W., Boyd C.D., Passmore H.C.;  
 RT "Use of an intron polymorphism to localize the tropoelastin gene to  
 RT mouse chromosome 7.";  
 RT chromosome 7.";  
 RL Genomics 23:125-131(1994).  
 CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND  
 CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
 CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
 CC INTO AN EXTENSIBLE 3D NETWORK.  
 CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U08210; AAA80155.1; -  
 DR HSSP; P04002; IWPA.  
 DR MGD; MGI:95317; Eln.  
 KW Structural protein; Repeat; Signal; Connective tissue.  
 FT SIGNAL 1  
 FT CHAIN 28 860 ELASTIN.  
 FT SEQUENCE 860 AA: 71955 MW: 0C0B5A8E1EDD7F1 CRC64;  
 SO

Query Match 10.4%; Score 177; DB 1; Length 860;  
 Best Local Similarity 27.4%; Pred. No. 0.03;  
 Matches 113; Conservative 20; Mismatches 185; Indels 94; Gaps 14;

19 AKAEEAASAA--SAATGNAAGDVVKNKSGAAAGK-----GEAASVNGIANGI 69  
 360 AAAAAAATAAATGAAAGLVPGGPRVRLPGAGIPGVGIGIPVGGIPGPGI 419  
 70 -----AAGKADAKKEGLDATTGAE--TTNVNAGKL--FVKRAADGDPADDA 116  
 420 GPGAVSPAAAAAATAAATAAAYGARGVGIPTYGVGAGCPGPGYGVGAGAGS 479  
 117 AAVAAASATGNAAGDVVNDVAKAKGDAASYNGIANGIKK-----IYDAEKADAK 169  
 480 AKAARYGAGGAGALGLVPGAVPGALPGAVPVPVPG-AGVPGAGTTPAAAAAATAA 538

```

Oy 170 E-----GKLNAAAGEHTNMDACKLFLYKNNKNGNYGEGGDA 204
Db 539 AGLPGVGCGVPGVGVGGIPGCVGVGVPCCVGGCGGTGIGAG-----PGJGAGACSPA 552
Oy 205 GKAAAAVAAYSGEOILKAIYHA-----AKDGEKCKKKAADRTNPIDAAIGGAGNDAA 258
Db 593 AAKSAKAKAAKQYRAAAGLGAAGVPGFGAGAGVPGFGAGAGVPGFGAGAGV 652
Oy 259 AAFATMKDDOIAAAMVLR-----GMAKDQGFALKDAAAAHBTGYKNA 301
Db 653 PGFGAGVPGSLAAKAAKYGAAAGLGGPGGLGGPGGLGGPGGL-----GGAGVPGRVAGA 708
Oy 302 VDIITKAAAEASASAAT-----GSAATGDIVVNGNGATFAKGDAAKSVYGIKNG 349
Db 709 AAPAAAAAATAAAKAAOYGLGAGGGLG-----AGGIGAGGLGAGGLGAGGLGAG 758

```

Search completed: January 10, 2002, 14:09:57  
Job time: 326 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:08:51 ; Search time 18.38 Seconds  
(without alignments)  
86.433 Million cell updates/sec

Title: US-09-445-803-14

Perfect score: 235  
Sequence: 1 PLTNPIDAIGSADRNMEA.....DQIAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 155098 seqs, 33800819 residues

1 number of hits satisfying chosen parameters: 155098

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: Pending\_Patents\_AA\_New:\*
- 2: /cgn2\_6/ptodata/1/paa/PCN\_NEW.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US06\_NEW.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US07\_NEW.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US08\_NEW.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US10\_NEW.COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US60\_NEW.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	54	23.0	198	5	US-09-708-427-66722
2	54	23.0	312	5	US-09-708-427-66721
3	54	23.0	394	5	US-09-708-427-66720
4	54	23.0	604	5	US-09-708-427-28541
5	54	23.0	620	5	US-09-708-427-28540
6	54	23.0	686	5	US-09-708-427-28539
7	53	22.6	359	5	US-09-897-516-5275
8	53	22.6	526	5	US-09-708-427-27506
9	53	22.6	544	5	US-09-708-427-27505
10	53	22.6	727	5	US-09-708-427-27504
11	52	22.1	1480	5	US-09-708-427-11856
12	52	22.1	1535	5	US-09-708-427-11855
13	52	22.1	1548	5	US-09-708-427-11854
14	51.5	21.9	114	5	US-09-154-707A-319
15	51.5	21.9	114	5	US-09-966-262-319
16	51	21.7	127	5	US-09-897-516-7117
17	51	21.7	127	5	US-09-611-526-2565
18	51	21.7	167	5	US-09-708-427-62406
19	51	21.7	201	5	US-09-708-427-62405
20	51	21.7	649	5	US-09-708-427-12799
21	50	21.3	174	5	US-09-708-427-65815
22	50	21.3	174	5	US-09-708-427-83240
23	50	21.3	423	5	US-09-708-427-70072
24	50	21.3	498	5	US-09-815-242-11777
25	50	21.3	585	5	US-09-708-427-25137
26	50	21.3	625	5	US-09-708-427-29600

27	50	21.3	626	5	US-09-708-427-25136	Sequence 25136, A
28	50	21.3	630	5	US-09-708-427-25135	Sequence 25135, A
29	50	21.3	648	5	US-09-468-646A-30	Sequence 30, Appl
30	50	21.3	648	5	US-09-468-646B-30	Sequence 30, Appl
31	50	21.3	675	5	US-09-708-427-29599	Sequence 29599, A
32	50	21.3	695	5	US-09-708-427-29598	Sequence 29598, A
33	49.5	21.1	177	5	US-09-708-427-76073	Sequence 76073, A
34	49.5	21.1	183	5	US-09-708-427-76072	Sequence 76072, A
35	49.5	21.1	193	5	US-09-708-427-76071	Sequence 76071, A
36	49.5	21.1	344	5	US-09-620-394B-966	Sequence 966, App
37	49.5	21.1	375	5	US-09-620-394B-965	Sequence 965, App
38	49.5	21.1	393	5	US-09-620-394B-964	Sequence 964, App
39	49	20.9	127	5	US-09-708-427-82550	Sequence 82550, A
40	49	20.9	139	5	US-09-708-427-62795	Sequence 62795, A
41	49	20.9	225	5	US-09-708-427-62794	Sequence 62794, A
42	49	20.9	262	5	US-09-708-427-62793	Sequence 62793, A
43	49	20.9	362	5	US-09-708-427-18950	Sequence 18950, A
44	49	20.9	371	5	US-09-708-427-18949	Sequence 18949, A
45	49	20.9	409	5	US-09-708-427-18948	Sequence 18948, A

ALIGNMENTS

RESULT 1  
US-09-708-427-66722  
Sequence 66722, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427.  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 66722  
LENGTH: 198  
TYPE: PRT  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..198  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc.feature  
LOCATION: 1..198  
OTHER INFORMATION: Ceres Seq. ID 1931147  
US-09-708-427-66722

Query Match 23.0%; Score 54; DB 5; Length 198;  
Best Local Similarity 27.9%; Pred. No. 6.8;  
Matches 12; Conservative 11; Mismatches 20; Indels 0; Gaps 0;

Qy 5 PIDAIGSADRNMEAEPDKMKDDQIAAMVLRGMAKDGQFAL 47  
Db 122 PLSCMNSCRPNAKAFKRDCKDHAVIATLRPISKDEITTI 164

RESULT 2  
US-09-708-427-66721  
Sequence 66721, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 66721  
LENGTH: 312



NAME/KEY: misc.feature  
LOCATION: 1..686  
OTHER INFORMATION: Ceres Seq. ID 1823500  
US-09-708-427-28539

Query Match 23.0%; Score 54; DB 5; Length 686;  
Best Local Similarity 36.6%; Pred. No. 31;  
Matches 15; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

OY 1 PLTNPIDAIG--GSADRNAEAFDKMKKDDQIAAAYLRGM 39  
Db 107 PLANTLVNVYKCGASHALGVFDEMPHRIHIAVSITAL 147

RESULT 7  
US-09-897-516-5275  
Sequence 5275, Application US/09897516  
GENERAL INFORMATION:  
APPLICANT: Cordin, David R.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Huesing, Joseph E.  
APPLICANT: Krasomil-Osterfeld, Karina C.  
APPLICANT: Malvar, Thomas M.  
APPLICANT: Slater, Steven C.  
APPLICANT: Spiridonov, Sergei  
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-21(51847)B  
CURRENT APPLICATION NUMBER: US/09/897,516  
CURRENT FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/215, 161  
PRIOR FILING DATE: 2000-06-30  
NUMBER OF SEQ ID NOS: 8409  
SEQ ID NO 5275  
LENGTH: 359  
TYPE: PRF  
ORGANISM: Xenorhabdus sp.  
US-09-897-516-5275

Query Match 22.6%; Score 53; DB 5; Length 359;  
Best Local Similarity 39.3%; Pred. No. 19;  
Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

OY 16 RNAEAFDKMKKDDQIAAAYLRGM 43  
Db 219 KNAKLYEAKKADGTGGYVRGDKDS 246

RESULT 8  
US-09-708-427-27506  
Sequence 27506, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 27506  
LENGTH: 526  
TYPE: PRF  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..526  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc.feature  
LOCATION: 1..526  
OTHER INFORMATION: Ceres Seq. ID 1821102

US-09-708-427-27506

Query Match 22.6%; Score 53; DB 5; Length 526;  
Best Local Similarity 29.7%; Pred. No. 30;  
Matches 11; Conservative 11; Mismatches 11; Indels 4; Gaps 1;

OY 12 GSADRNAEAFDKMKD---DQIAAAYLRGMADQ 44  
Db 111 GNLDSLVRWDEMRDEIKPVMAYGLVGLCKDGR 147

RESULT 9  
US-09-708-427-27505  
Sequence 27505, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 27505  
LENGTH: 544  
TYPE: PRF  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..544  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc.feature  
LOCATION: 1..544  
OTHER INFORMATION: Ceres Seq. ID 1821101  
US-09-708-427-27505

Query Match 22.6%; Score 53; DB 5; Length 544;  
Best Local Similarity 29.7%; Pred. No. 32;  
Matches 11; Conservative 11; Mismatches 11; Indels 4; Gaps 1;

OY 12 GSADRNAEAFDKMKD---DQIAAAYLRGMADQ 44  
Db 129 GNLDSLVRWDEMRDEIKPVMAYGLVGLCKDGR 165

RESULT 10  
US-09-708-427-27504  
Sequence 27504, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 27504  
LENGTH: 727  
TYPE: PRF  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..727  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc.feature  
LOCATION: 1..727  
OTHER INFORMATION: Ceres Seq. ID 1821100  
US-09-708-427-27504

Query Match 22.6%: Score 53; DB 5; Length 727;  
Best Local Similarity 29.7%: Pred. NO. 45;  
Matches 11; Conservative 11; Mismatches 11; Indels 4; Gaps 1;  
QY 12 GSADRNAAEPKMKD---DQIAAAMVLRGMARDQ 44  
DB 312 GNLDAJSLRWDEMRDRDEKPDVMAVGTIVGLCKDGR 348

RESULT 11  
US-09-708-427-11856  
; Sequence 11856, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11856  
; LENGTH: 1480  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1480  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1480  
; OTHER INFORMATION: Ceres Seq. ID 1823914  
US-09-708-427-11856

Query Match 22.1%: Score 52; DB 5; Length 1480;  
Best Local Similarity 38.5%: Pred. NO. 1,4e+02;  
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
QY 19 EAFDKMKDDQIAAAMVLRGMARDQ 44  
DB 109 KVFDMPEKDLVANNVINGFAENGK 134

RESULT 12  
US-09-708-427-11855  
; Sequence 11855, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11855  
; LENGTH: 1535  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1535  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1535  
; OTHER INFORMATION: Ceres Seq. ID 1823913  
US-09-708-427-11855

Query Match 22.1%: Score 52; DB 5; Length 1535;  
Best Local Similarity 38.5%: Pred. NO. 1,5e+02;  
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 19 EAFDKMKDDQIAAAMVLRGMARDQ 44  
DB 164 KVFDMPEKDLVANNVINGFAENGK 189

RESULT 13  
US-09-708-427-11854  
; Sequence 11854, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11854  
; LENGTH: 1548  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1548  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1548  
; OTHER INFORMATION: Ceres Seq. ID 1823912  
US-09-708-427-11854

Query Match 22.1%: Score 52; DB 5; Length 1548;  
Best Local Similarity 38.5%: Pred. NO. 1,5e+02;  
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
QY 19 EAFDKMKDDQIAAAMVLRGMARDQ 44  
DB 177 KVFDMPEKDLVANNVINGFAENGK 202

RESULT 14  
US-09-154-707A-319  
; Sequence 319, Application US/09154707A  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: P2004P1  
; CURRENT APPLICATION NUMBER: US/09/154,707A  
; CURRENT FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: PCT/US98/05311  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: US 60/041,277  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/042,344  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/041,276  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/041,281  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/048,094  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,350  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,188  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,135  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/050,937  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,187  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: US 60/048,099

PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,352  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,186  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,069  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,095  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,131  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,096  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,355  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,160  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,351  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,154  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/054,804  
 PRIOR FILING DATE: 1997-08-05  
 PRIOR APPLICATION NUMBER: US 60/056,370  
 PRIOR FILING DATE: 1997-08-19  
 PRIOR APPLICATION NUMBER: US 60/060,862  
 PRIOR FILING DATE: 1997-10-02  
 NUMBER OF SEQ ID NOS: 343  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 319  
 LENGTH: 114  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: MISC\_FEATURE  
 LOCATION: (2)  
 OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally  
 NAME/KEY: MISC\_FEATURE  
 LOCATION: (114)  
 OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally  
 US-09-154-707A-319

Query Match 21.9% Score 51.5; DB 5; Length 114;  
 Best Local Similarity 34.1%; Pred. No. 7.5;  
 Matches 14; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

5 PIDAIGSSADRNNAEPDKMKKDDQIAAM--VLRGMKAD 42  
 9 PVEAAGAGADGPREPASERASRAEPPAVAMGNDLMGTAED 49

RESULT 15  
 US-09-966-262-319  
 Sequence 319, Application US/09966262  
 GENERAL INFORMATION:  
 APPLICANT: Young et al.  
 TITLE OF INVENTION: 87 Human Secreted Proteins  
 FILE REFERENCE: P2004P1  
 CURRENT APPLICATION NUMBER: US 09/154,707  
 PRIOR APPLICATION NUMBER: US 09/154,707  
 PRIOR FILING DATE: 1998-09-17  
 PRIOR APPLICATION NUMBER: PCT/US98/05311  
 PRIOR FILING DATE: 1998-03-19  
 PRIOR APPLICATION NUMBER: US 60/041,277  
 PRIOR FILING DATE: 1997-03-21  
 PRIOR APPLICATION NUMBER: US 60/042,344  
 PRIOR FILING DATE: 1997-03-21  
 PRIOR APPLICATION NUMBER: US 60/041,276  
 PRIOR FILING DATE: 1997-03-21  
 PRIOR APPLICATION NUMBER: US 60/041,281  
 PRIOR FILING DATE: 1997-03-21

PRIOR APPLICATION NUMBER: US 60/048,094  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,350  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,188  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,135  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/050,937  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,187  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,099  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,352  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,186  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,069  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,095  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,131  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,096  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,355  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,160  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/048,351  
 PRIOR FILING DATE: 1997-05-30  
 PRIOR APPLICATION NUMBER: US 60/054,804  
 PRIOR FILING DATE: 1997-08-05  
 PRIOR APPLICATION NUMBER: US 60/056,370  
 PRIOR FILING DATE: 1997-08-19  
 PRIOR APPLICATION NUMBER: US 60/060,862  
 PRIOR FILING DATE: 1997-10-02  
 NUMBER OF SEQ ID NOS: 343  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 319  
 LENGTH: 114  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: MISC\_FEATURE  
 LOCATION: (2)  
 OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally  
 NAME/KEY: MISC\_FEATURE  
 LOCATION: (114)  
 OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally  
 US-09-966-262-319

Query Match 21.9% Score 51.5; DB 5; Length 114;  
 Best Local Similarity 34.1%; Pred. No. 7.5; Matches 14; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

5 PIDAIGSSADRNNAEPDKMKKDDQIAAM--VLRGMKAD 42  
 9 PVEAAGAGADGPREPASERASRAEPPAVAMGNDLMGTAED 49

Search completed: January 10, 2002, 14:08:52  
 Job time: 321 sec

Thu Jan 10 14:14:50 2002

us-09-445-803-14.rapn

Page 6

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:03:56 ; Search time 19.72 Seconds  
(Without alignments)  
53.634 Million cell updates/sec

Title: US-09-445-803-14

Sequence: 1 PLTNPIDAIGSGSADRMAA.....DOIAAAYLRGMADGQFAL 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Database : Issued Patents AA: \* 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

1: /cgn2\_6/prodata/2/iaa/5A COMB pep: \*  
2: /cgn2\_6/prodata/2/iaa/5B COMB pep: \*  
3: /cgn2\_6/prodata/2/iaa/5A COMB pep: \*  
4: /cgn2\_6/prodata/2/iaa/5B COMB pep: \*  
5: /cgn2\_6/prodata/2/iaa/5A COMB pep: \*  
6: /cgn2\_6/prodata/2/iaa/5B COMB pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50.5	21.5	1253	1	Sequence 12, Appl
2	50.5	21.5	1261	1	Sequence 18, Appl
3	49	20.9	553	2	Sequence 36, Appl
4	49	20.9	553	3	Sequence 36, Appl
5	48.5	20.6	344	1	Sequence 24, Appl
6	48	20.4	170	4	Sequence 50, Appl
7	47	20.0	388	2	Sequence 56, Appl
8	47	20.0	486	2	Sequence 3, Appl
9	47	20.0	524	2	Sequence 2, Appl
10	47	20.0	803	4	Sequence 2, Appl
11	47	20.0	1455	2	Sequence 2, Appl
12	47	20.0	1618	1	Sequence 4, Appl
13	47	20.0	3033	1	Sequence 8, Appl
14	47	20.0	3033	1	Sequence 9, Appl
15	46.5	19.8	1338	1	Sequence 9, Appl
16	46.5	19.8	1338	2	Sequence 9, Appl
17	46.5	19.8	1529	4	Sequence 10, Appl
18	46.5	19.8	1529	4	Sequence 10, Appl
19	46.5	19.8	1599	2	Sequence 9, Appl
20	46.5	19.8	1600	2	Sequence 10, Appl
21	46	19.6	381	1	Sequence 6, Appl
22	46	19.6	4551	3	Sequence 1, Appl
23	46	19.6	4613	4	Sequence 31, Appl
24	46	19.6	11877	4	Sequence 6, Appl
25	45.5	19.4	382	3	Sequence 7, Appl
26	45.5	19.4	402	4	Sequence 22, Appl
27	45.5	19.4	645	4	Sequence 2, Appl

28	45.5	19.4	3782	4	US-09-105-537-4	Sequence 4, Appl
29	45	19.1	338	1	US-08-210-394-1	Sequence 1, Appl
30	45	19.1	506	1	US-08-369-780-2	Sequence 2, Appl
31	45	19.1	506	1	US-08-475-682-2	Sequence 2, Appl
32	45	19.1	506	1	US-08-780-833-2	Sequence 2, Appl
33	45	19.1	506	1	US-08-636-036-2	Sequence 2, Appl
34	45	19.1	506	3	US-08-918-509-2	Sequence 2, Appl
35	45	19.1	506	3	US-09-108-262-2	Sequence 2, Appl
36	45	19.1	578	1	US-08-458-120-2	Sequence 2, Appl
37	45	19.1	578	2	US-08-867-970-2	Sequence 2, Appl
38	45	19.1	578	2	US-09-326-217-2	Sequence 2, Appl
39	44.5	18.9	318	3	US-08-484-322-50	Sequence 50, Appl
40	44.5	18.9	318	3	US-08-671-947-2	Sequence 2, Appl
41	44.5	18.9	861	4	US-08-960-048-12	Sequence 12, Appl
42	44	18.7	35	4	US-09-082-279B-939	Sequence 939, App
43	44	18.7	35	4	US-09-082-279B-940	Sequence 940, App
44	44	18.7	486	2	US-08-942-423-2	Sequence 2, Appl
45	44	18.7	706	1	US-08-339-152A-29	Sequence 29, Appl

#### ALIGNMENTS

RESULT 1  
US-08-252-966B-12  
Sequence 12, Application US/08252966B  
Patent No. 5624818  
GENERAL INFORMATION:  
APPLICANT: Eisenman, Robert N.  
APPLICANT: Hurlin, Peter J.  
TITLE OF INVENTION: Regulatory Proteins that Dimerize with  
NUMBER OF INVENTION: Mad or Max  
CORRESPONDENCE ADDRESSES: 19  
ADDRESS: Christensen, O'Connor, Johnson, and KindnessPULC  
STREET: 1420 Fifth Ave., Suite 2800  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/252,966B  
FILING DATE: 01-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Shelton, Dennis K.  
REGISTRATION NUMBER: 26,997.  
REFERENCE/DOCKET NUMBER: FHCRI7694  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 682-8100  
TELEFAX: (206) 224-0779  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: translation of mslna cDNA; see Figure 23  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
US-08-252-966B-12  
Query Match 21.5%; Score 50.5; DB 1; Length 1253;  
Best Local Similarity 26.7%; Pred. No. 58;







INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
STRANDEDNESS: both  
TOPOLOGY: both  
US-08-282-197C-56

Query Match 20.0%; Score 47; DB 2; Length 388;  
Best Local Similarity 37.0%; Pred. No. 44;  
Matches 10; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Oy 14 ADRNAEAFDKMKDDQIAAAAVLRGMA 40  
Db 320 AKRYQELFDALKENMDIVSAVFWGIS 346

## RESULT 8

US-942-423-3  
Sequence 3, Application US/08942423  
Patent No. 5891673

## GENERAL INFORMATION:

APPLICANT: Hashimoto, Yasuhiro  
APPLICANT: Takemoto, Yoshihiro  
TITLE OF INVENTION: Lck Binding Protein  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syntex (U.S.A.) Inc.  
STREET: 3401 Hillview Ave.  
CITY: Palo Alto  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94303

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,423  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,715  
FILING DATE: 23-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Perles, Rohan

REGISTRATION NUMBER: 35,752  
REFERENCE/DOCKET NUMBER: 28260  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 852-1698  
TELEFAX: (415) 496-3529

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: LCK BINDING PROTEIN  
US-08-942-423-3

Query Match 20.0%; Score 47; DB 2; Length 486;  
Best Local Similarity 33.3%; Pred. No. 58;  
Matches 13; Conservative 8; Mismatches 14; Indels 4; Gaps 1;

Oy 3 TNPDAIGGSADRNAL-----AFDKMKDDQIAAAVLR 37

Db 225 TTPIEAASSGARGLKAFESMAEERKREDEKAAQVAVR 263

## RESULT 9

US-08-615-942A-2  
Sequence 2, Application US/08615942A  
Patent No. 5863532

## GENERAL INFORMATION:

APPLICANT: JOLINDA A. TRAUH  
APPLICANT: REGINA D. ROONEY  
APPLICANT: ROLF JAKOBI  
APPLICANT: POLYGENA T. TUAZON  
APPLICANT: CHARNG-JUI CHEN  
APPLICANT: WILLIAM E. MEER  
APPLICANT: EDWARD J. CARROLL, JR.

TITLE OF INVENTION: Compositions and Methods Comprising  
TITLE OF INVENTION: Cytostatic Protein Kinase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Belliner & Carson  
STREET: 201 N. Figueroa St., Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90012

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/615,942A  
FILING DATE:

## CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: NONE  
FILING DATE: NONE  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean K.  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 1279-203XX

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213/977-1001  
TELEFAX: 213/977-1003

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 524 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
US-08-615-942A-2

Query Match 20.0%; Score 47; DB 2; Length 524;  
Best Local Similarity 39.4%; Pred. No. 64;  
Matches 13; Conservative 6; Mismatches 8; Indels 6; Gaps 2;

Oy 4 NPIDAIGGS-ADRNAEAFDKMK-----DDQI 30  
Db 200 DPIDPVGDSHVSQAKSPDKKKTKMTDEI 232

## RESULT 10

US-09-063-035-2  
Sequence 2, Application US/09063035  
Patent No. 6160091

## GENERAL INFORMATION:



OY 1 PLTNPIDAIGSADRNAAFDKMKKDDQ 29  
Db 690 PLTKENQEPRLSRLEDEKKEAFRLSEKQ 718

## RESULT 13

US-07-925-695-8  
; Sequence 8, Application US/07925695  
; Patent No. 5428145  
; GENERAL INFORMATION:  
; APPLICANT: OKAMOTO, Hiroaki  
; APPLICANT: NAKAMURA, Tetsuo  
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,  
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND  
; NUMBER OF INVENTION: DETECTION SYSTEMS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Beveridge, Degrandt, Wellacher & Young  
; STREET: 1850 M Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/07/925.695  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: JP 287402/91  
; FILING DATE: 09-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 360441/91  
; FILING DATE: 05-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wellacher, Robert G.  
; REGISTRATION NUMBER: 20,531  
; REFERENCE/DOCKET NUMBER: 06/87-48009  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 659-1462  
; TELEFAX: (202) 659-1462  
; TELEX: WU1 64470  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3033 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: linear  
US-07-925-695-8

Query Match 20.0%; Score 47; DB 1; Length 3033;  
Best Local Similarity 22.2%; Pred. No. 5.8e+02;  
Matches 18; Conservative 7; Mismatches 20; Indels 36; Gaps 2;

OY 1 PLTNPIDAIGSAD-----RNAEAFDKMKKDDQ----- 29  
Db 1119 PGTSLDPTCTCGAAYDLVYTRNADYIPVRRKDDRCALLSPRLSTLKSGSGPYLCSRG 1178

OY 30 -----IAAAWVLRGMARDQGF 45

Db 1179 HAVGLFRAAVCARGVAKSIDF 1199

## RESULT 14

US-07-925-695-9  
; Sequence 9, Application US/07925695  
; Patent No. 5428145

; GENERAL INFORMATION:  
; APPLICANT: OKAMOTO, Hiroaki  
; APPLICANT: NAKAMURA, Tetsuo  
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,  
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND  
; NUMBER OF INVENTION: DETECTION SYSTEMS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Beveridge, Degrandt, Wellacher & Young  
; STREET: 1850 M Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/925.695  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: JP 287402/91  
; FILING DATE: 09-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 360441/91  
; FILING DATE: 05-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wellacher, Robert G.  
; REGISTRATION NUMBER: 20,531  
; REFERENCE/DOCKET NUMBER: 06/87-48009  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 659-1462  
; TELEFAX: (202) 659-1462  
; TELEX: WU1 64470  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3033 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: linear  
US-07-925-695-9

Query Match 20.0%; Score 47; DB 1; Length 3033;  
Best Local Similarity 22.2%; Pred. No. 5.8e+02;  
Matches 18; Conservative 7; Mismatches 20; Indels 36; Gaps 2;

OY 1 PLTNPIDAIGSAD-----RNAEAFDKMKKDDQ----- 29  
Db 1119 PGTSLDPTCTCGAAYDLVYTRNADYIPVRRKDDRCALLSPRLSTLKSGSGPYLCSRG 1178

OY 30 -----IAAAWVLRGMARDQGF 45

Db 1179 HAVGLFRAAVCARGVAKSIDF 1199

## RESULT 15

US-08-728-470-9  
; Sequence 9, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: OF NO. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:05:14 : Search time 40.15 Seconds  
(without alignments)  
86.711 Million cell updates/sec

Title: US-09-445-803-14

Sequence: 1 PLTNPIDAIGSGADPMAEA.....DOIAMAVLRGMKDGQFAL 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Minimum DB seq length: 0 522463  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_1101.\*  
2: /SIDSR/gcgdata/geneseq/AA1980.DAT.\*  
3: /SIDSR/gcgdata/geneseq/AA1981.DAT.\*  
4: /SIDSR/gcgdata/geneseq/AA1982.DAT.\*  
5: /SIDSR/gcgdata/geneseq/AA1983.DAT.\*  
6: /SIDSR/gcgdata/geneseq/AA1984.DAT.\*  
7: /SIDSR/gcgdata/geneseq/AA1985.DAT.\*  
8: /SIDSR/gcgdata/geneseq/AA1986.DAT.\*  
9: /SIDSR/gcgdata/geneseq/AA1987.DAT.\*  
10: /SIDSR/gcgdata/geneseq/AA1988.DAT.\*  
11: /SIDSR/gcgdata/geneseq/AA1989.DAT.\*  
12: /SIDSR/gcgdata/geneseq/AA1990.DAT.\*  
13: /SIDSR/gcgdata/geneseq/AA1992.DAT.\*  
14: /SIDSR/gcgdata/geneseq/AA1993.DAT.\*  
15: /SIDSR/gcgdata/geneseq/AA1994.DAT.\*  
16: /SIDSR/gcgdata/geneseq/AA1995.DAT.\*  
17: /SIDSR/gcgdata/geneseq/AA1996.DAT.\*  
18: /SIDSR/gcgdata/geneseq/AA1997.DAT.\*  
19: /SIDSR/gcgdata/geneseq/AA1998.DAT.\*  
20: /SIDSR/gcgdata/geneseq/AA1999.DAT.\*  
21: /SIDSR/gcgdata/geneseq/AA2000.DAT.\*  
22: /SIDSR/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235	100.0	47	20	AAW95613
2	176.5	75.1	349	20	AAW95612
3	176.5	75.1	349	21	AAAB36281
4	144.5	61.5	1328	20	AAW20088
5	142	60.4	156	20	AAW20113
6	142	60.4	533	20	AAW20112
7	131.5	56.0	168	20	AAW20089
8	130	55.3	356	18	AAW22676
9	117	49.8	26	21	AAAB36273
10	117	49.8	26	21	AAAB36280
11	55	23.4	593	22	AAW91309

12	52.5	22.3	354	22	AAW91164	C glutamicum prote
13	51.5	21.9	245	20	AAW21884	Pha acetocetyl-Co
14	51	21.7	127	21	AAW7333	HTM clone 1760185
15	51	21.7	383	19	AAW48336	Brucella abortus 3
16	50.5	21.5	969	21	AAAB01827	Haemophilus influe
17	50.5	21.5	975	21	AAAB01826	Haemophilus influe
18	50.5	21.5	1026	22	AAW94315	Human protein sequ
19	50.5	21.5	1253	18	AAW10038	Mad binding protei
20	50.5	21.5	1261	18	AAW10040	Mad binding protei
21	49.5	21.1	143	21	AAW45798	Arabidopsis thalia
22	49.5	21.1	143	21	AAW45798	Arabidopsis thalia
23	49.5	21.1	177	21	AAW33768	Arabidopsis thalia
24	49.5	21.1	183	21	AAW33767	Arabidopsis thalia
25	49.5	21.1	185	21	AAW17255	Arabidopsis thalia
26	49.5	21.1	185	21	AAW45797	Arabidopsis thalia
27	49.5	21.1	185	21	AAW45825	Arabidopsis thalia
28	49.5	21.1	193	21	AAW33766	Arabidopsis thalia
29	49.5	21.1	206	21	AAW17254	Arabidopsis thalia
30	49.5	21.1	206	21	AAW22849	Arabidopsis thalia
31	49.5	21.1	206	21	AAW45796	Arabidopsis thalia
32	49.5	21.1	206	21	AAW45825	Arabidopsis thalia
33	49.5	21.1	257	21	AAW17253	Arabidopsis thalia
34	49.5	21.1	257	21	AAW22848	Arabidopsis thalia
35	49.5	21.1	287	21	AAW22847	Arabidopsis thalia
36	49.5	21.1	297	22	AAW1486	S. epidermidis ope
37	49.5	21.1	297	22	AAW1826	S. epidermidis ope
38	49.5	21.1	311	21	AAW404819	Arabidopsis thalia
39	49.5	21.1	317	21	AAW404818	Arabidopsis thalia
40	49.5	21.1	319	22	AAW22875	S. epidermidis ope
41	49.5	21.1	324	21	AAW15081	Arabidopsis thalia
42	49.5	21.1	327	21	AAW404817	Arabidopsis thalia
43	49.5	21.1	344	21	AAW13282	Arabidopsis thalia
44	49.5	21.1	375	21	AAW13281	Arabidopsis thalia
45	49.5	21.1	393	21	AAW13280	Arabidopsis thalia

## ALIGNMENTS

RESULT 1	
AAW95613	standard; Protein: 47 AA.
ID	AAW95613
XX	
AC	AAW95613
XX	
DT	08-JUN-1999 (first entry)
XX	
DE	Borrelia burgdorferi surface antigen p39.5 clone 14-1 polypeptide.
KW	Lyme disease; surface antigen; p39.5; diagnosis; prevention;
KW	vaccine; antisense; therapy; treatment; primer; probe;
KW	antibody; DNA.
XX	
OS	Borrelia burgdorferi.
XX	
FT	
FT	key
FT	CDS
FT	Location/Qualifiers
FT	2..142
FT	/*tag= a
FT	/product= "clone 14-1 encoded protein"
XX	
PN	W09900413-A1.
XX	
PD	07-JAN-1999.
XX	
PF	29-JUN-1998; 98WO-US13551.
XX	
PR	30-JUN-1997; 97US-0051271.
XX	
PA	(TULA ) TULANE EDUCATIONAL FUND.
XX	
PI	Phillip MT;
XX	
DR	WPI; 1999-095676/08.

DR N-PSDB; AAX07422.  
XX  
PT New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi  
PT - and related vectors, transformants, antibodies and polypeptides,  
PT for diagnosis, prevention and treatment of Lyme disease  
XX  
PS Claim 5; Page 60; 89pp; English.  
XX  
CC The sequence is that of a Borrelia burgdorferi surface antigen P39.5  
CC clone 14-1 polypeptide. It can be used in the production of P39.5 or  
CC fragments of it which may be used to raise antibodies to, and in the  
CC development of vaccines against Lyme disease. The sequence can  
CC also be used for making primers and probes for diagnosis, also in  
CC DNA vaccines, as antisense therapeutics and for drug screening.  
CC Antibodies can be used as diagnostic (immunoassay) reagents, for  
CC treating Lyme disease, for affinity purification, for drug  
CC screening and to produce anti-idiotypic antibodies (used in the  
CC same way as P39.5 to induce an immune response).  
XX

Sequence 47 AA:

Query Match 100.0%; Score 235; DB 20; Length 47;  
Best Local Similarity 100.0%; Pred. No. 2.4e-26;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTTNPDAATGGSADRN-AEAFDKMKKDDQIAAAMVLRGMKDGQFAL 47  
Db 1 tptnpdaatggsadrdnaaefdkmkddqiaaamvltirmakdgqfal 47

#### RESULT 2

AAM95612  
ID AAM95612 standard; Protein; 349 AA.  
XX

AAM95612;  
XX

DT 08-JUN-1999 (first entry)  
XX

DE Borrelia burgdorferi surface antigen P39.5 clone 7-1 polypeptide.  
XX

XX Lyme disease; surface antigen; P39.5; diagnosis; prevention;  
KW vaccine; antisense; therapy; treatment; primer; probe;  
XX antibody; DNA.

XX Borrelia burgdorferi.  
OS

XX  
PN W09900413-A1.  
XX

XX 07-JAN-1999.  
XX

XX 29-JUN-1998; 98WO-US13551.  
PF

XX 30-JUN-1997; 97US-0051271.  
PR

XX (TULANE ) TULANE EDUCATIONAL FUND.  
PA

XX PhilIpp MT;  
PI

XX WPI; 1999-095676/08.  
XX

XX N-PSDB; AAX07411.  
DR

XX New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi  
PT - and related vectors, transformants, antibodies and polypeptides,  
PT for diagnosis, prevention and treatment of Lyme disease

XX  
PS Claim 5; Pages 54-55; 89pp; English.  
XX

XX The sequence is that of a Borrelia burgdorferi surface antigen P39.5  
CC clone 7-1 polypeptide. It can be used in the production of P39.5  
CC or fragments of it which may be used to raise antibodies to, and  
CC in the development of vaccines against Lyme disease. The sequence  
CC can also be used for making primers and probes for diagnosis, also

CC in DNA vaccines, as antisense therapeutics and for drug screening.  
CC Antibodies can be used as diagnostic (immunoassay) reagents, for  
CC treating Lyme disease, for affinity purification, for drug  
CC screening and to produce anti-idiotypic antibodies (used in the  
CC same way as P39.5 to induce an immune response).  
XX

SO Sequence 349 AA:

Query Match 75.1%; Score 176.5; DB 20; Length 349;  
Best Local Similarity 82.6%; Pred. No. 5.5e-17;  
Matches 38; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 3 TNPDAATGGSADRN-AEAFDKMKKDDQIAAAMVLRGMKDGQFAL 47  
Db 242 tnpdaatggsadrdnaaefdkmkddqiaaamvltirmakdgqfal 287

#### RESULT 3

AAB36281  
ID AAB36281 standard; peptide; 349 AA.  
XX

AC AAB36281;  
XX

DT 21-FEB-2001 (first entry)  
XX

DE B. garinii P7-1 protein.  
XX

XX Variable surface antigen; invariable region; VISE; Lyme disease;  
KW Lyme borreliosis.

XX Borrelia garinii.  
OS

XX W0200065064-A1.  
PN

XX 02-NOV-2000.  
PD

XX 25-APR-2000; 2000WO-US11085.  
PF

XX 28-APR-1999; 99US-0300971.  
PR

XX (TULANE ) TULANE EDUCATIONAL FUND.  
PA

XX PhilIpp MT, Llang FT;  
PI

XX WPI; 2000-687350/67.  
DR

XX Novel peptides comprising an invariable 26-amino acid long region  
PT isolated from Borrelia burgdorferi (sensu lato), useful for rapid and  
PT specific diagnosis of Lyme disease -

XX  
PS Example 2; Fig 2; 76pp; English.  
XX

XX The present invention describes several peptides comprised of the  
CC invariable regions IR1-IR6 of the B. burgdorferi variable surface antigen  
CC (VISE) variable domain. These peptides can be used in the specific  
CC diagnosis, treatment and vaccination against B. burgdorferi, which causes  
CC Lyme disease (also known as Lyme borreliosis) in humans, dogs, horses,  
CC cows and other animals.  
XX

XX  
SO Sequence 349 AA:

Query Match 75.1%; Score 176.5; DB 21; Length 349;  
Best Local Similarity 82.6%; Pred. No. 5.5e-17;  
Matches 38; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 3 TNPDAATGGSADRN-AEAFDKMKKDDQIAAAMVLRGMKDGQFAL 47  
Db 242 tnpdaatggsadrdnaaefdkmkddqiaaamvltirmakdgqfal 287

#### RESULT 4



ID	AAV20088		standard; Protein: 1328 AA.
AC	AAV20088;		
DT	19-JUL-1999	(first entry)	
DE	B. burgdorferi antigenic protein, f24-1.aa.		
KW	Antigenic protein; vaccine; Lyme disease; infection; detection.		
OS	Borrelia burgdorferi.		
PN	W09859071-A1.		
PD	30-DEC-1998.		
PE	18-JUN-1998;	98WO-USJ2718.	
P	03-SEP-1997;	97US-0057483.	
P	10-JUN-1997;	97US-0050359.	
PR	22-JUL-1987;	97US-0053344.	
XX	22-JUL-1997;	97US-0053377.	
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	(MEDI-) MEDIMUNE INC.		
F1	Choi GH, Erwin AL, Hanson MS, Lathigra R;		
DR	WPI: 1999-189980/16.		
DR	N-PSDB; AAX61785.		
PT	New isolated Borrelia burgdorferi nucleic acids - used to develop products for the diagnosis, prevention and treatment of diseases caused by Borrelia, particularly Lyme disease		
PS	Claim 12; Page 192-193; 275pp; English.		
CC	This sequence represents a Borrelia burgdorferi (Bb) protein of the invention, which is suitable for use in a vaccine. The Bb polypeptides can be used in vaccines for eliciting protective antibodies to members of the Borrelia genus, particularly for the use against Lyme disease in humans and animals. They can be used for preventing or attenuating an infection caused by a member of the Borrelia genus. The products can also be used for detection of members of the Borrelia genus.		
SQ	Sequence	1328 AA;	
SC	Score	61.5%;	
SC	Best Local Similarity	71.1%;	
SC	Matches	32; Conservative	5; Mismatches
SC			7; Indels
SC			1; Gaps
OY	4 NPIDAIG-GSADRNAEAFDKMKKDDQIAAVALWGMAKGQFAL	47	
DB	::    ::       ::       ::		
DB	479 npiaaaiqknadgdafgdmkddqaaataatrgmakdgrfav	523	
RESULT	5		
ID	AAV20113		
AC	AAV20113 standard; Protein: 156 AA.		
DT	19-JUL-1999	(first entry)	
DE	B. burgdorferi antigenic protein, t49-2.aa.		
KW	Antigenic protein; vaccine; Lyme disease; infection; detection.		
OS	Borrelia burgdorferi.		
PN	W09859071-A1		

XX	30-DEC-1998.	
PD		
XX	18-JUN-1998;	98WO-US12718.
PF		
XX	03-SEP-1997;	97US-0057483.
PR	20-JUN-1997;	97US-0050359.
PR	22-JUL-1997;	97US-0053344.
PR	22-JUL-1997;	97US-0053377.
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	(MEDI-) MEDIMMUNE INC.	
XX		
PI	Choi GH, Erwin AL, Hanson MS, Lathigra R;	
XX		
DR	WPI: 1999-189980/16.	
XX	N-PSDB: AAX61810.	
PT	New isolated Borrelia burgdorferi nucleic acids - used to develop	
PT	products for the diagnosis, prevention and treatment of diseases	
PT	caused by Borrelia, particularly Lyme disease	
XX		
PS	Claim 12: Page 202; 275pp; English.	
XX		
CC	This sequence represents a Borrelia burgdorferi (Bb) protein of the	
CC	invention, which is suitable for use in a vaccine. The Bb polypeptides	
CC	can be used in vaccines for eliciting protective antibodies to members of	
CC	the Borrelia genus, particularly for the use against Lyme disease in	
CC	humans and animals. They can be used for preventing or attenuating an	
CC	infection caused by a member of the Borrelia genus. The products can also	
CC	be used for detection of members of the Borrelia genus.	
XX		
SO	Sequence	156 AA:
	Query Match	60.4%; Score 142; DB 20; Length 156;
	Best Local Similarity	68.2%; Pred. No. 1,7e-12;
	Matches	30; Conservative; 5; Mismatches
		9; Indels
		0; Gaps
		0;
OY	4 NEIDAIGGSADRNAEAFDKMKRDDQIAAAVLRGMADGQFAL	47
	:    :      :	
Db	100 nplaaajgknedgaeffkdemkkddqiaaatalrgmaxdgkftav	143
RESULT	6	
AAAY20112		
ID	AAAY20112 standard; Protein: 533 AA.	
XX		
AC	AAAY20112;	
XX		
DT	19-JUL-1999 (first entry)	
XX		
DE	B. burgdorferi antigenic protein, f49-2.aa.	
XX		
XX	Antigenic protein; vaccine; Lyme disease; infection; detection.	
OS	Borrelia burgdorferi.	
XX		
PN	MO9859071-AL.	
XX		
PD	30-DEC-1998.	
XX		
PF	18-JUN-1998;	98WO-US12718.
XX		
PR	03-SEP-1997;	97US-0057483.
PR	20-JUN-1997;	97US-0050359.
PR	22-JUL-1997;	97US-0053344.
PR	22-JUL-1997;	97US-0053377.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	(MEDI-) MEDIMMUNE INC.	
PI	Choi GH, Erwin AL, Hanson MS, Lathigra R;	





PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI: 2001-376931/40.  
DR N-PSDB: AAH66528.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
XX Claim 17; SEQ ID NO: 5063; 246bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 593 AA:  
SQ  
Query Match 23.4%; Score 55; DB 22; Length 593;  
Best Local Similarity 41.4%; Pred. No. 21;  
Matches 12; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
QY 13 SADRRAEFDFKMKDDQIAAAVLRGMKAK 41  
Db 356 sldissealdkldedakagasmmaakk 384  
RESULT 12  
AAG91164  
ID AAG91164 standard; Protein: 354 AA.  
XX  
XX AAG91164;  
A  
DT 26-SEP-2001 (first entry)  
XX  
XX C glutamicum protein fragment SEQ ID NO: 4918.  
DE  
XX  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.  
XX  
XX Corynebacterium glutamicum.  
OS  
XX  
XX EPI108790-A2.  
PN  
XX  
XX 20-JUN-2001.  
PD  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
PF  
XX  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
PT

XX  
XX WPI: 2001-376931/40.  
DR N-PSDB: AAH66383.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
XX Claim 17; SEQ ID NO: 4918; 246bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 354 AA:  
SQ  
Query Match 22.3%; Score 52.5; DB 22; Length 354;  
Best Local Similarity 32.6%; Pred. No. 25;  
Matches 14; Conservative 8; Mismatches 20; Indels 1; Gaps 1;  
QY 4 NPIDALIGSADRRAEFDFKMKDDQIAAAVLRGMKDGQA 46  
Db 74 nnelqvwsvgygvesteslkddidtaevpfihlva-gela 115  
RESULT 13  
AA921894  
ID AA921894 standard; Protein: 245 AA.  
XX  
XX AA921894;  
AC  
XX  
XX 24-SEP-1999 (first entry)  
DT  
XX  
XX PHA acetoacetyl-CoA reductase sequence.  
DE  
XX  
XX Polyhydroxyalkanoate; PHA; PHA biosynthesis; PHA synthase; plastic;  
KW PHA beta-ketothiolase; PHA acetoacetyl-CoA reductase; Alcaligenes latus;  
KW biodegradable polymer.  
XX  
XX Alcaligenes latus.  
OS  
XX  
XX WO9936547-A1.  
PN  
XX  
XX 22-JUL-1999.  
PD  
XX  
XX 19-JAN-1999; 99WO-KR00031.  
PF  
XX  
XX 26-DEC-1998; 98KR-0058760.  
PR 19-JAN-1998; 98KR-0001422.  
PR 19-JAN-1998; 98KR-0001423.  
XX  
XX (GLDS ) LG CHEM LTD.  
PA  
XX  
XX Choi J, Choo S, Han K, Hong S, Huh T, Lee SY;  
PI Lee Y, Song J, Yoon H;  
XX  
XX WPI: 1999-444403/37.  
DR N-PSDB: AAX86965, AAX86968.  
XX  
XX New polyhydroxyalkanoate biosynthesis-related genes used in the  
PT production of biodegradable polymers



CC to detect and/or quantify the cellular immune response in an  
CC individual, or detect individuals who have been in contact with  
CC Brucella.  
xx  
SQ Sequence 383 AA;

Query Match	21.7%	Score 51	DB 19	Length 383
Best Local Similarity	34.1%	Pred. No. 46		
Matches	15	Conservative	7	Mismatches 8
				Indels 14
				Gaps 2
Oy	9	AIGSGADNRNAEAFDMMK-----	DDQIAAAYLRGMA	40
		:	: :	
Db	185	alg--sdcmkgafdmkslryvdnfgicrnlsamwlgxa		226

Search completed: January 10, 2002, 14:05:15  
Job time: 169 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:08:25 ; Search time 184.82 Seconds  
(without alignments)  
70.609 Million cell updates/sec

Title: US-09-445-803-14

Perfect score: 235  
Sequence: 1 PLTNPIDAIGSADRNAA.....DQIAAAWLRGMKDGQFAL 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235	100.0	47	18	US-09-445-803-14
2	176.5	75.1	349	17	US-09-300-971A-9
3	176.5	75.1	349	18	US-09-445-803-2
4	164.5	70.0	189	17	US-09-300-971A-10
5	148.5	63.2	179	15	US-09-125-619-18
6	147.5	62.8	216	15	US-09-125-619-38
7	146.5	62.3	197	15	US-09-125-619-25
8	145.5	61.9	216	15	US-09-125-619-46
9	144.5	61.5	190	15	US-09-125-619-19

10	144.5	61.5	323	17	US-09-300-971A-11	Sequence 11, Appl
11	144	61.3	189	15	US-09-125-619-26	Sequence 26, Appl
12	143.5	61.1	212	15	US-09-125-619-32	Sequence 32, Appl
13	143.5	61.1	212	15	US-09-125-619-44	Sequence 44, Appl
14	142.5	60.6	215	15	US-09-125-619-36	Sequence 16, Appl
15	142	60.6	169	15	US-09-125-619-16	Sequence 41, Appl
16	142	60.6	212	15	US-09-125-619-41	Sequence 23, Appl
17	141.5	60.2	187	15	US-09-125-619-43	Sequence 45, Appl
18	141.5	60.2	209	15	US-09-125-619-45	Sequence 27, Appl
19	140.5	59.8	190	15	US-09-125-619-27	Sequence 28, Appl
20	139.5	59.4	212	15	US-09-125-619-48	Sequence 24, Appl
21	139	59.1	194	15	US-09-125-619-28	Sequence 39, Appl
22	138	58.7	214	15	US-09-125-619-34	Sequence 17, Appl
23	136.5	58.1	169	15	US-09-125-619-24	Sequence 35, Appl
24	136.5	58.1	213	15	US-09-125-619-39	Sequence 47, Appl
25	136	57.9	195	15	US-09-125-619-37	Sequence 40, Appl
26	136	57.9	214	15	US-09-125-619-35	Sequence 22, Appl
27	134.5	57.2	208	15	US-09-125-619-47	Sequence 33, Appl
28	133.5	56.8	209	15	US-09-125-619-42	Sequence 15, Appl
29	133.5	56.8	211	15	US-09-125-619-40	Sequence 43, Appl
30	132	56.2	211	15	US-09-125-619-40	Sequence 2, Appl1
31	131.5	56.0	169	15	US-09-125-619-22	Sequence 13, Appl
32	131.5	56.0	190	15	US-09-125-619-22	Sequence 21, Appl
33	131	55.7	209	15	US-09-125-619-33	Sequence 20, Appl
34	130	55.3	190	15	US-09-125-619-15	Sequence 8, Appl1
35	130	55.3	212	15	US-09-125-619-31	Sequence 1, Appl1
36	130	55.3	212	15	US-09-125-619-43	Sequence 8, Appl1
37	130	55.3	356	15	PCT-US97-02952-2	
38	130	55.3	356	15	US-09-125-619-2	
39	130	55.3	356	15	US-09-125-619-13	
40	128	54.5	189	15	US-09-125-619-21	
41	126.5	53.8	189	15	US-09-125-619-20	
42	117	49.8	26	17	US-09-300-971-1	
43	117	49.8	26	17	US-09-300-971-8	
44	117	49.8	26	17	US-09-300-971A-1	
45	117	49.8	26	17	US-09-300-971A-8	

## ALIGNMENTS

RESULT 1  
US-09-445-803-14  
Sequence 14, Appl1  
GENERAL INFORMATION:  
APPLICANT: Adminis. of Tulane Educational, Fund  
TITLE OF INVENTION: Surface Antigens and Proteins Useful in  
Compositions for the Diagnosis and Prevention of Lyme Disease  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/445,803  
FILING DATE: 13-Dec-1999  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/051,271  
FILING DATE: 30-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: TUL2APCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 47 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-445-803-14

Query Match 100.0%; Score 235; DB 18; Length 47;  
Best Local Similarity 100.0%; Pred. No. 1.3e-25;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 1 PLTNPIDAIGSADRNAEAFDKMKDDQIAAAYLRGMKADGOFAL 47  
1 PLTNPIDAIGSADRNAEAFDKMKDDQIAAAYLRGMKADGOFAL 47

RESULT 2  
US-09-300-971A-9  
Sequence 9, Application US/09300971A  
GENERAL INFORMATION:  
APPLICANT: Philipp, Mario T.  
APPLICANT: Liang, Fang Ting  
TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease  
FILE REFERENCE: TUL3USA  
CURRENT APPLICATION NUMBER: US/09/300.971A  
CURRENT FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 9  
LENGTH: 349  
TYPE: PRT  
ORGANISM: Borrelia garinii  
US-09-300-971A-9

Query Match 75.1%; Score 176.5; DB 17; Length 349;  
Best Local Similarity 82.6%; Pred. No. 4.1e-16;  
Matches 38; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Query 3 TNPIDAIGGSADRN-AEAFDKMKDDQIAAAYLRGMKADGOFAL 47  
242 TNPIDAIGGSADRN-AEAFDKMKDDQIAAAYLRGMKADGOFAL 287

RESULT 3  
US-09-445-803-2  
Sequence 2, Application US/09445803  
GENERAL INFORMATION:  
APPLICANT: Adminis. of Tulane Educational, Fund  
APPLICANT: Philipp, Mario T.  
TITLE OF INVENTION: Surface Antigens and Proteins Useful in  
Diagnosis for the Diagnosis and Prevention of Lyme  
Disease  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/445.803  
FILING DATE: 13-Dec-1999  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/051,271  
FILING DATE: 30-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: TUL2APCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-445-803-2

Query Match 75.1%; Score 176.5; DB 18; Length 349;  
Best Local Similarity 82.6%; Pred. No. 4.1e-16;  
Matches 38; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Query 3 TNPIDAIGGSADRN-AEAFDKMKDDQIAAAYLRGMKADGOFAL 47  
Db 242 TNPIDAIGGSADRN-AEAFDKMKDDQIAAAYLRGMKADGOFAL 287

RESULT 4  
US-09-300-971A-10  
Sequence 10, Application US/09300971A  
GENERAL INFORMATION:  
APPLICANT: Philipp, Mario T.  
APPLICANT: Liang, Fang Ting  
TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease  
FILE REFERENCE: TUL3USA  
CURRENT APPLICATION NUMBER: US/09/300.971A  
CURRENT FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 10  
LENGTH: 189  
TYPE: PRT  
ORGANISM: Borrelia burgdorferi  
US-09-300-971A-10

Query Match 70.0%; Score 164.5; DB 17; Length 189;  
Best Local Similarity 73.9%; Pred. No. 9.1e-15;  
Matches 34; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

Query 3 TNPIDAIGGSADRNAEAF-DKMKDDQIAAAYLRGMKADGOFAL 47  
Db 128 TNPIDAIGSTGDNDAAAFQDEMKKNDQIAAAYLRGMKADGOFAL 173

RESULT 5  
US-09-125-619-18  
Sequence 18, Application US/09125619  
GENERAL INFORMATION:  
APPLICANT: NORRIS, STEVEN J.  
APPLICANT: JING-REN, ZHANG  
APPLICANT: HARDHAM, JOHN M.  
APPLICANT: HOWELL, JERRILYN K.  
APPLICANT: BARBOUR, ALAN G.  
APPLICANT: WEINSTOCK, GEORGE M.  
TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA  
FILE REFERENCE: UTSH:234









**THIS PAGE BLANK (USPTO)**

[illegible]



Db 61 EAAGAAGEANEDAAKLFKAANNANANGGAGAPDAEKAANAASVSGKQILKATVDAAGKE 120  
 QY 220 GGEKGGKAAADRTNPTDAIGAGNDAAAFATMKKDDQIAAAYLVKMAKDGOFALKD 289  
 Db 121 --EKKGAKEAENPTIAAIGSTGKER--AAAFSHMKKNDQIAAAYLVKMAKDGOFALKD 176  
 QY 290 AAA 292  
 Db 177 ADA 179

## RESULT 5

Q9RHW4 PRELIMINARY: PRT: 190 AA.  
 AC Q9RHW4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VLSF15 PROTEIN (FRAGMENT).  
 C VLSF15  
 OC Borrelia burgdorferi (Lyme disease spirochete).  
 OC Plasmid 20-kb borrelial plasmid.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=297;  
 RX MEDLINE=98181050; PubMed=9514637;  
 RA Kawabata H., Myouga F., Inagaki Y., Murali N., Watanabe H.;  
 "Genetic and immunological analyses of Vls (Vmp-like sequences) of  
 RT Borrelia burgdorferi.";  
 RL Microb. Pathog. 24:155-166(1998).  
 DR EMBL, AB011066; BA87889.1;  
 DR InterPro: IPR000680; Borrelia\_lipo.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 KW Plasmid.  
 FT NON\_TER 1 1  
 FT 190 190  
 SO SEQUENCE 190 AA; 18748 MW; D55EA566DA6D29C8 CRC64;

Query Match 26.5%; Score 452.5; DB 2; Length 190;  
 Best Local Similarity 62.2%; Pred. No. 1,6e-14;  
 Matches 112; Conservative 18; Mismatches 41; Indels 9; Gaps 7;

QY 120 AASAATGNA--AIGDV--VNGDVAKAGGDAASVNGIAKIGIYDAEKADEKGLNA 175  
 Db 3 AATAASGKEDMIGKVKYNAAGAAGGEEITSVNGIASGIGIYTAAEKA--GEEGKLS 61  
 176 -AGAGCTTNADAGKLFVNKAGNVGEGADGAGKAAAVAAVSGEOLIKATVHAAKDGGEKQ 234  
 Db 62 EAAGGGEAENEDGKLFPAK-NDTGGDAKAEKAAAVSAVSGKQILKATVDAKGDGGERK 120  
 QY 235 G-KKAADRTNPTDAIGAGNDAAA-AFATMKKDDQIAAAYLVKMAKDGOFALKDAAA 292  
 Db 121 GVDVDYKKAENPTIAAIGSTGNDAAAFSKDEMKKNDQVAAAYLVKMAKDGOFALKNADA 180

## RESULT 6

Q9RF40 PRELIMINARY: PRT: 224 AA.  
 AC Q9RF40;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VLSF (FRAGMENT).  
 GN VLSF.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W12M;

RA Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;  
 "Conservation and heterogeneity of vlsE among human and tick isolates  
 RT of Borrelia burgdorferi.";  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL, AF201679; AAF25655.1;  
 DR InterPro: IPR000680; Borrelia\_lipo.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 DR NON\_TER 1 1  
 DR 224 224  
 FT NON\_TER 1 1  
 FT 224 224  
 SO SEQUENCE 224 AA; 21807 MW; FB0900B83615E7B CRC64;

Query Match 25.6%; Score 438; DB 2; Length 224;  
 Best Local Similarity 50.4%; Pred. No. 9e-14;  
 Matches 120; Conservative 26; Mismatches 68; Indels 24; Gaps 9;

QY 79 GRLDATTAGCTTNVNAAGKLFVRAADGGDADAGKAAAVAAVSAATGNAIGDVNGDV 138  
 Db 4 GKPDSTGSGVTA-----VEGAIKEVSELLDKLVKAVKTAEGASSGTAIGEYV-ADA 54  
 QY 139 AKAAGDAASVNGIAKIGIYDAEKADEKGLNA-AGAECTTAAAGKLFVNAGNV 197  
 Db 55 GAAKVAADKASVTVGIAKIGIYEAAGSE---KKAVAAGAESKKEAGKLFKAGANN 110  
 QY 198 GGEAGDAGKAAAVAAVSGEOLIKATVHAADG-GEKGGKKAADRTNPTDAIGAGDND 256  
 Db 111 AGDSEAKSAAGASVAVSGEOLISALVTAGAGAEQBEKKRPADAKNP1AAI--GAGDEE 169  
 QY 257 AAAAFA--TMKDDQIAAAYLVKMAKDGOFALKDAAAHEGTVKNAVDTIRKAAEAA 312  
 Db 170 DCAEFEGKDEMKKDDQIAAAYLVKMAKDGKFAVKD---DEKRAEAGA---IKGAESA 221

## RESULT 7

Q06878 PRELIMINARY: PRT: 356 AA.  
 AC Q06878;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VMP-Like VLSF.  
 GN VLSF.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B31 (ATCC 35210);  
 RX MEDLINE=97262068; PubMed=9108482;  
 RA Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;  
 "Antigenic variation in Lyme disease borreliae by promiscuous  
 RT recombination of Vmp-like sequence cassettes.";  
 RL Cell 89:275-285(1997).  
 DR EMBL, U76405; AAC4573.1;  
 DR InterPro: IPR000680; Borrelia\_lipo.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 SO SEQUENCE 356 AA; 36012 MW; 304035731B2AA2BE CRC64;

Query Match 25.6%; Score 437.5; DB 2; Length 356;  
 Best Local Similarity 46.9%; Pred. No. 1.5e-13;  
 Matches 127; Conservative 26; Mismatches 87; Indels 31; Gaps 11;

QY 79 GRLDATTAGCTTNVNAAGKLFVRAADGGDADAGKAAAVAAVSAATGNAIGDVNGDV 138  
 Db 103 GKPDSTGSGVTA-----VEGAIKEVSELLDKLVKAVKTAEGASSGTAIGEYV-ADA 153  
 QY 139 AKAAGDAASVNGIAKIGIYDAEKADEKGLNA-AGAECTTAAAGKLFVNAGNV 197  
 Db 154 DAAKVAADKASVTVGIAKIGIYEAAGSE---KKAVAAGAESKKEAGKLFKAGAA 209  
 QY 198 GGEAGDAGKAAAVAAVSGEOLIKATVHAADGGEKGGKKAADRTNPTDAIGAGDND 257

DB 210 HGDSEASAKAGAVSAVSGEQILISAIVTA--DAEODGKKPEAKNPITAAI---GDKDG 265  
OY 258 AAFPA--TWKKDDQIAAAMVLRGMKADGOFALKDAAAHEGVKNAVDITKAAFAA--S 313  
DB 266 GAEFODEKKDDQIAAAILALRGMAKDKGFAYKD-----GEKEAEGATIKGAESAAR 319  
OY 314 AASATGSAATIGDVVNGMGTAKGDAKSVN 344  
DB 320 VLGAITG--LIGDAVS--SGLRKVGDVSKAAS 347

RESULT 8  
O68375 PRELIMINARY: PRT: 217 AA.

AC 068375; PRELIMINARY: PRT: 217 AA.  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DR VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).  
GN VSLE.  
OC Borrelia burgdorferi (Lyme disease Spirochete).  
OG Plasmid lp28-1.  
OX Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OC NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B31-1500B.  
RA Zhang J.-R., Norris S.J.;  
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF034523; AAC31352.1; .  
DR InterPro: IPR000680; Borrelia\_lipo.  
DR Pfam: PF00921; Lipoprotein\_2; 1.  
KW Plasmid.  
FT NON\_TER  
KW NON\_TER  
SQ SEQUENCE 217 AA; 21109 MW; 7834DC72AA30D83D CRC64;

Query Match 25.4%; Score 434.5; DB 2; Length 217;  
Best Local Similarity 49.6%; Pred. No. 1.3e-13;  
Matches 115; Conservative 22; Mismatches 68; Indels 27; Gaps 7;

OY 79 GKLDTAGAGTTNNVNAAGKLFVRRAADGGDADGAKAAVAASAATGNAATGVDVNGDV 138  
DB 4 GKPDSGVS GTA-----VEGAIKEVSELDKLVKAVKTAEGASSGTDATGKVVNDNA 55  
OY 139 AAKAGDAASVNGIAKIGIYDAAEKADAKGKLN--AGAGTTNADAGKLFVNAGNV 197  
DB 56 GAAKAAKRESVNGIAKIGIYEAAGSE---KLKVAATGCKNNKRAKGLFGK----- 106  
OY 198 GGEAGDAG-----KAAAVAAVSGEQILISAIVTAHAKDGEKOGKKAADRTNPIDAAIGG 251  
DB 107 AGDADGADSEASAKAGAVSAVSGEQILISAIVKAA--DAGDDEGKKPADAATNPIDAAIGK 165  
OY 252 AGDNDAAAAFAATMKDDQIAAAMVLRGMKADGOFALKDA--AAAHGTYKNA 301  
DB 166 GNEDEAFENQDKKDDQIAAAILALRGMAKDKGFAYKDDDEKGAEGAIGKA 217

RESULT 9  
O07055 PRELIMINARY: PRT: 358 AA.  
AC 007055; PRELIMINARY: PRT: 358 AA.  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DR VMP-LIKE SEQUENCE PROTEIN VSLE.  
GN VSLE.  
OC Borrelia burgdorferi (Lyme disease Spirochete).  
OG Plasmid lp28-1.  
OX Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OC NCBI\_TaxID=139;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-B31-MIEAC;  
RX MEDLINE-97262068; PubMed-9108482;  
RA Zhang J.-R., Hardham J.M., Barbour A.G., Norris S.J.;  
RT "Antigenic variation of VMP-like sequence cassettes."  
RL Cell 89:275-285(1997).  
DR EMBL: U84556; AAC45764.1; .  
DR InterPro: IPR000680; Borrelia\_lipo.  
DR Pfam: PF00921; Lipoprotein\_2; 1.  
KW Plasmid.  
SQ SEQUENCE 358 AA; 36009 MW; 6022AE9MAA076EEC CRC64;

Query Match 25.4%; Score 433.5; DB 2; Length 358;  
Best Local Similarity 46.4%; Pred. No. 2.4e-13;  
Matches 128; Conservative 26; Mismatches 83; Indels 39; Gaps 12;

OY 79 GKLDTAGAGTTNNVNAAGKLFVRRAADGGDADGAKAAVAASAATGNAATGVDVNGDV 138  
DB 103 GKPDSGVS GTA-----VEGAIKEVSELDKLVKAVKTAEGASSGTAATGGEV-ADA 153  
OY 139 AAKAGDAASVNGIAKIGIYDAAEKADAKGKLNAGAGETTNADAGKLFVNAGNV 198  
DB 154 AKA--ADKDSVNGIAKIGIYEAAGSE---KLKVAATGESNKGAKLF-----GKVD 203  
OY 199 GEAGD---AGKAAVAVSAVSGEQILISAIVTAHAKDGEKOGKKAADRTNPIDAAIGGA 252  
DB 204 GAAGDSEASAKAGAVSAVSGEQILISAIVTAAGAAASEADGCKRVADTFNPIDAAIGG 263  
OY 253 GNDAAAFAATMKDDQIAAAMVLRGMKADGOFALK--DAAAHEGVKNAVDITKAAAE 310  
DB 264 NEENGAEFGDGKMKDDQIAAAILALRGMAKDKGFAYKNDDEKKAEGA-----IKGAE 316  
OY 311 AA--SAASATGSAATIGDVVNGMGTAKGDAKSVN 344  
DB 317 SAVRKVLGAITG--LIGDAVS--SGLRKVGDVSKAAS 349

RESULT 10  
O68361 PRELIMINARY: PRT: 212 AA.  
AC 068361; PRELIMINARY: PRT: 212 AA.  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DR VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).  
GN VSLE.  
OC Borrelia burgdorferi (Lyme disease Spirochete).  
OG Plasmid lp28-1.  
OX Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OC NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B31-1394D;  
RA Zhang J.-R., Norris S.J.;  
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF034509; AAC31338.1; .  
DR InterPro: IPR000680; Borrelia\_lipo.  
DR Pfam: PF00921; Lipoprotein\_2; 1.  
KW Plasmid.  
FT NON\_TER  
KW NON\_TER  
SQ SEQUENCE 212 AA; 20488 MW; A98950CCE798BDF CRC64;

Query Match 25.1%; Score 429; DB 2; Length 212;  
Best Local Similarity 50.0%; Pred. No. 2.2e-13;  
Matches 113; Conservative 24; Mismatches 69; Indels 20; Gaps 7;

OY 79 GKLDTAGAGTTNNVNAAGKLFVRRAADGGDADGAKAAVAASAATGNAATGVDVNGDV 138  
DB 4 GKPDSGVS GTA-----VEGAIKEVSELDKLVKAVKTAEGASSGTAATGGEVNDND- 54



QY 139 AKAKGDAASVNGIAKIGIYDAAEKADAKEGKLNAGAEAGTTNADGKLFYKNAAGV 198  
 DB 55 --AKVADKASVNGIAKIGIYEIVAAAGSE----KLKVAATGEMNKAGKLGKAGADAH 108  
 QY 199 GEAGDAGKAAAVAAVSGEOLIKATVHAARKDGEKGRKAADRTNPIDAIGAGDNDAA 258  
 DB 109 GSEASAKRAAGAVASVSGEOLISATYTA--DAAEQDGKPEAKNPITAAI--CKGDGAG 166  
 QY 259 AAFAT--MKDDQIAAAMVLRGNAKDGOFALK--DAAAHEGYKNA 301  
 DB 167 ADFGDMKKDDQIAAALRGMAKDGKFAVKNNEKGAEGAIKGA 212

## RESULT 11

ID 068368 PRELIMINARY: PRT: 212 AA.

AC 068368;  
 DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
 DE VMP-LIKE SEQUENCE PROTEIN VLSL (FRAGMENT).  
 GN VLSL.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Plasmid lp28-1.  
 OC Bacteria: Spirochaetales; Spirochaetaceae; Borrelia.  
 NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B31-1395F;  
 RA Zhang J.-R., Norris S.J.;  
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF034516; AAC31345.1; -;  
 DR InterPro: IPR000680; Borrelia\_11po.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 KW Plasmid.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 212 AA: 20645 MW; C7DAE7377957B175 CRC64;

Query Match 25.1%; Score 428; DB 2; Length 212;  
 Best Local Similarity 50.4%; Pred. No. 2.4e-13;  
 Matches 114; Conservative 20; Mismatches 72; Indels 20; Gaps 7;

QY 79 GKLDATGAGCTTNVNAAGKLFVKRAADGGDADAGKAAAVAAASATGNAATGVDVNGDV 138  
 DB 4 GKPSTGSGVTA-----VEGAIKEVSELLDKLVKAVKTAGACSSGTAIGEVV-ADA 54  
 DB 139 AKAKGDAASVNGIAKIGIYDAAEKADAKEGKLNAGAEAGTTNADGKLFYKNAAGV 198  
 DB 55 DAAKADKASVNGIAKIGIYEIVAAAGSE---KLKAAAEENNNKAGKLGKAGADAH 110  
 QY 199 GEAGDAGKAAAVAAVSGEOLIKATVHAARKDGEKGRKAADRTNPIDAIGAGDNDAA 258  
 DB 111 GSEASAKRAAGAVASVSGEOLISATYTA--DAAEQDGKPEAKNPITAAI--GKDGAG 166  
 QY 259 AAFAT--TMKKDDQIAAAMVLRGNAKDGOFALKDAAAAN--EGTYKNA 301  
 DB 167 ADFGDMKKDDQIAAALRGMAKDGKFAVKNNEKGAEGAIKGA 212

## RESULT 12

ID 09RF44 PRELIMINARY: PRT: 222 AA.

AC 09RF44;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
 DE VLSL (FRAGMENT).  
 GN Borrelia burgdorferi (Lyme disease spirochete).  
 OS Plasmid 32 kb.

OC Bacteria: Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B294;  
 RA Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;  
 RT "Conservation and heterogeneity of vlsE among human and tick isolates of Borrelia burgdorferi."  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF201675; AAF25651.1; -;  
 DR InterPro: IPR000680; Borrelia\_11po.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 KW Plasmid.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 222 AA: 21635 MW; 6E4782BB9A092ACF CRC64;

Query Match 24.9%; Score 426; DB 2; Length 222;  
 Best Local Similarity 50.0%; Pred. No. 3.2e-13;  
 Matches 115; Conservative 21; Mismatches 76; Indels 18; Gaps 7;

QY 78 ECKLDATGAGCTTNVNAAGKLFVKRAADGGDADAGKAAAVAAASATGNAATGVDVNGDV 137  
 DB 2 DGPSTGSGVTA-----VEGAIKEVSELLDKLVKAVKTAGAPSGTDAIGEVV-AD 52  
 QY 138 VAKAKGDAASVNGIAKIGIYDAAEKADAKEGKLNAGAEAGTTNADGKLFYKNAAGV 197  
 DB 53 AGAKADKASVNGIAKIGIYEIVAAAGSE---KLKAAAEENNNKAGKLGKAGADAH 108  
 QY 198 GEAGDAGKAAAVAAVSGEOLIKATVHAARKDGEKGRKAADRTNPIDAIGAGDND 256  
 DB 109 GSEASAKRAAGAVASVSGEOLISATYTA--DAAEQDGKPEAKNPITAAI--GKGDAD 167  
 QY 257 AAFAT--ATMKDDQIAAAMVLRGNAKDGOFALK--DAAAHEGYKNAVD 303  
 DB 168 AGADFHEMKKDDQIAAALRGMAKDGKFAVKNDEKGAEGAIKGA 217

## RESULT 13

ID 09RF43 PRELIMINARY: PRT: 223 AA.

AC 09RF43;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
 DE VLSL (FRAGMENT).  
 GN VLSL.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Plasmid 32 kb.  
 OC Bacteria: Spirochaetales; Spirochaetaceae; Borrelia.  
 NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B294;  
 RA Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;  
 RT "Conservation and heterogeneity of vlsE among human and tick isolates of Borrelia burgdorferi."  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF201676; AAF25652.1; -;  
 DR InterPro: IPR000680; Borrelia\_11po.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 KW Plasmid.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 223 AA: 21636 MW; 0B344689C0DC3A0 CRC64;

Query Match 24.9%; Score 426; DB 2; Length 223;  
 Best Local Similarity 50.2%; Pred. No. 3.2e-13;  
 Matches 115; Conservative 21; Mismatches 75; Indels 18; Gaps 7;

QY 79 GKLDATGAGCTTNVNAAGKLFVKRAADGGDADAGKAAAVAAASATGNAATGVDVNGDV 138

```

Db      4 GKPDSSTGVSCTA-----VEGAIKEVSELDKLVKAKYKTEGASSGDAIGEVV-ADA 54
OY      139 ARAKGGDAASVNGIAKIGIYDAAEKADAKGKLNAGAEGTTNDAKGLFVKNNGVNG 198
Db      55 GAAKADAKASVTGIAKIGIYEIYAAGGSE----KLKAAAAGENNKKAKKLFKVDAAHA 110
OY      199 GEAGDAGKAAAANAASVSGEQLIKAIYHAAKDG-GEKQKKAADRTNPIDAAIGAGDND 257
Db      111 GDSEASAKAAGAVASVSGEQLISAIYKAAAGAEODGSEKPADAKNPIDAAI-GRGDADA 169
OY      258 AAFA-ATMKKDDQIAAAMVLKGMARDGOFALK--DAAAHEGTVKNA 303
Db      170 GADEFHEMKKDDQIAAAILRGMARDGKFAVKNDEKGRAGAIGKA 218

RESULT 14
ID      007056 PRELIMINARY: PRT: 209 AA.
007056 01-JUL-1997 (Tremblrel. 04, Created)
01-JUL-1997 (Tremblrel. 04, Last sequence update)
01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE VMP-LIKE SEQUENCE PROTEIN VLSE (FRAGMENT).
GN VLSE.
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp28-1.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId-139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B31-M1EAD;
RA MEDIUM-97262068; PubMed-9108482;
RH Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT "Antigenic variation in Lyme disease borreliae by promiscuous
recombination of VMP-like sequence cassettes.";
RL Cell 89:275-285(1997).
DR EMBL: U84557; AAC5194.1;
DR InterPro: IPR000680; Borrelia_lp28-1.
DR Pfam: PF00921; Lipoprotein_2; 1.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 209 209
SQ SEQUENCE 209 AA: 20214 MW: 5350AE7F83B7B38C CRC64;

Query Match 24.8%; Score 423.5; DB 2: Length 209;
Best Local Similarity 49.3%; Pred. No. 3.9e-13;
Matches 111; Conservative 24; Mismatches 69; Indels 21; Gaps 7;

OY 79 GKIDATGAGCTTNVNAAGKLFVKRAADGGDADAGKAAAANAASATGNAIGDVVNGDV 138
Db 4 GKPDSSTGVSCTA-----VEGAIKEVSELDKLVTAVKTAEGASSGTDALIGEYVND- 54
OY 139 ARAKGGDAASVNGIAKIGIYDAAEKADAKGKLNAGAEGTTNDAKGLFVKNNGVNG 198
Db 55 --AKADKASVYGIAGIYEIYAAGGSE---KLKVAAGKESNKGAKLFGKAGADAN 108
OY 199 GEAGDAGKAAAANAASVSGEQLIKAIYHAAKDGEGEKQKKAADRTNPIDAAIGAGDND 257
Db 109 GDSEASAKAAGAVASVSGEQLISAIYTA--DAAEODGSEKPADAKNPIDAAI-GRGDADF 167
OY 258 AAFAATMKKDDQIAAAMVLKGMARDGOFALK--DAAAHEGTVKNA 301
Db 168 GDC---MKKDDQIAAAILRGMARDGKFAVKNDEKKAEGATKGA 209

RESULT 15
ID      068358 PRELIMINARY: PRT: 214 AA.
AC      068358;
DT      01-AUG-1998 (Tremblrel. 07, Created)
DT      01-AUG-1998 (Tremblrel. 07, Last sequence update)

```

```

DT      01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE VMP-LIKE SEQUENCE PROTEIN VLSE (FRAGMENT).
GN VLSE.
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp28-1.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId-139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B31-1380C;
RA Zhang J.-R., Norris S.J.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF034506; AAC31335.1;
DR InterPro: IPR000680; Borrelia_lp28-1.
DR Pfam: PF00921; Lipoprotein_2; 1.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 214 214
SQ SEQUENCE 214 AA: 20857 MW: 12B38DEB1A457277 CRC64;

```

```

Query Match 24.7%; Score 422; DB 2: Length 214;
Best Local Similarity 49.8%; Pred. No. 4.7e-13;
Matches 113; Conservative 21; Mismatches 73; Indels 20; Gaps 7;

OY 79 GKIDATGAGCTTNVNAAGKLFVKRAADGGDADAGKAAAANAASATGNAIGDVVNGDV 138
Db 4 GKPDSSTGVSCTA-----VEGAIKEVSELDKLVKAVTAEGASSGTDALIGEYVANDA 55
OY 139 ARAKGGDAASVNGIAKIGIYDAAEKADAKGKLNAGAEGTTNDAKGLFVKNNGVN 197
Db 56 DAAKVAADKASVNGIAKIGIYEIYAAGGSE---KLKAAVAAKESNKGAKLFGKAGAAA 111
OY 198 GEAGDAGKAAAANAASVSGEQLIKAIYHAAKDGEGEKQKKAADRTNPIDAAIGAGDND 257
Db 112 HDSEASAKAAGAVASVSGEQLISAIYTA--DAAEODGKPEAKNPIDAAI--GDKGS 167
OY 258 AAFA--TMKDDQIAAAMVLKGMARDGOFALKDAAA--EGTVKNA 301
Db 168 GAFFGODEMKKDDQIAAAILRGMARDGKFAVKNDEKKEKAEGATKGA 214

```

Search completed: January 10, 2002, 14:09:34  
Job time: 333 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:09:34 ; Search time 38 Seconds  
(without alignments)  
180.916 Million cell updates/sec

Title: US-09-445-803-14

Perfect score: 235

Sequence: 1 PUTNPTDAIGSADRNARF.....DQIAAMVLRGMAKDGCFAL 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164	69.8	192	2	Q9RHWS
2	164	69.8	223	2	Q9RHWS
3	154	65.5	190	2	Q9RHWS
4	149.5	63.6	211	2	Q9RHWS
5	149.5	63.6	218	2	Q9RHWS
6	148.5	63.2	192	2	Q9RHWS
7	147.5	62.8	217	2	Q9RHWS
8	145.5	61.9	358	2	Q9RHWS
9	144.5	61.5	190	2	Q9RHWS
10	144.5	61.5	211	2	Q9RHWS
11	144.5	61.5	212	2	Q9RHWS
12	144.5	61.5	213	2	Q9RHWS
13	144.5	61.5	213	2	Q9RHWS
14	144	61.3	189	2	Q9RHWS
15	143.5	61.1	215	2	Q9RHWS
16	143.5	61.1	216	2	Q9RHWS
17	143.5	61.1	216	2	Q9RHWS
18	143.5	61.1	356	2	Q9RHWS
19	142.5	60.6	215	2	Q9RHWS

20	142	60.4	169	2	Q9RHWS	006879	borrelia bu
21	142	60.4	210	2	Q9RHWS	007064	borrelia bu
22	141.5	60.2	186	2	Q9RHWS	006886	borrelia bu
23	141.5	60.2	189	2	Q9RHWS	007054	borrelia bu
24	141.5	60.2	210	2	Q9RHWS	007054	borrelia bu
25	141.5	60.2	211	2	Q9RHWS	0068301	borrelia bu
26	141.5	60.2	212	2	Q9RHWS	0068341	borrelia bu
27	141.5	60.2	212	2	Q9RHWS	0068344	borrelia bu
28	141.5	60.2	212	2	Q9RHWS	0068360	borrelia bu
29	141.5	60.2	214	2	Q9RHWS	0068302	borrelia bu
30	141.5	60.2	218	2	Q9RHWS	0068353	borrelia bu
31	140.5	59.8	190	2	Q9RHWS	006889	borrelia bu
32	140.5	59.8	212	2	Q9RHWS	0068361	borrelia bu
33	140.5	59.8	213	2	Q9RHWS	0068373	borrelia bu
34	140.5	59.8	214	2	Q9RHWS	0068362	borrelia bu
35	140.5	59.8	222	2	Q9RHWS	0068362	borrelia bu
36	140.5	59.8	223	2	Q9RHWS	0068372	borrelia bu
37	139.5	59.4	212	2	Q9RHWS	0068372	borrelia bu
38	139.5	59.4	216	2	Q9RHWS	0068355	borrelia bu
39	139	59.1	176	2	Q9RHWS	006890	borrelia bu
40	139	59.1	215	2	Q9RHWS	0068364	borrelia bu
41	139	59.1	216	2	Q9RHWS	0068365	borrelia bu
42	139	59.1	217	2	Q9RHWS	0068304	borrelia bu
43	139	59.1	221	2	Q9RHWS	0068305	borrelia bu
44	138	58.7	211	2	Q9RHWS	0068305	borrelia bu
45	138	58.7	214	2	Q9RHWS	007058	borrelia bu

#### ALIGNMENTS

RESULT 1  
Q9RHWS PRELIMINARY: PRT: 192 AA.  
AC Q9RHWS  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VLSF8 PROTEIN (FRAGMENT).  
GN VLSF8.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Borrelia 20-kb borrelial plasmid.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=297;  
RX MEDLINE=98181050; PubMed=9514637;  
RA Kawabata H., Myounga F., Inagaki Y., Mural N., Watanabe H.;  
RT "Genetic and immunological analyses of Vls (VMP-like sequences) of  
RT Borrelia burgdorferi.";  
RT Microb. Pathog. 24:155-166(1998).  
DR EMBL: AB011065; BAA87888.1; -.  
DR InterPro: IPR00680; Borrelia\_lipo.  
DR Pfam: PF00921; Lipoprotein\_2; 1.  
KW Plasmid.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 192 AA; 18638 MW; F45DEIDF40A81A12 CRC64;

Query Match 69.8%; Score 164; DB 2; Length 192;  
Best Local Similarity 72.3%; Pred. No. 2.3e-13;  
Matches 34; Conservative 7; Mismatches 4; Indels 2; Gaps 1;

QY 3 TNPDAIGSADRNARF--DKMKDDQIAAMVLRGMAKDGCFAL 47  
DB 131 TNPDAIGSADRNARF--DKMKDDQIAAMVLRGMAKDGCFAL 177

RESULT 2  
Q9RHWS PRELIMINARY: PRT: 223 AA.



	RESULT	6	
	006881		
ID	006881	PRELIMINARY;	PRT; 192 AA.
AC	006881;		
DT	01-JUL-1997 (TREMblrel. 04, Created)		
DT	01-JUL-1997 (TREMblrel. 04, Last sequence update)		
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)		
DE	PUTATIVE VLS RECOMBINATION CASSETTE VLISA (FRAGMENT).		
OS	Borrelia burgdorferi ( Lyme disease spirochete ).		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.		
OX	NCBI_TaxID=139;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-B31 (ATCC 35210) CLONE 5A3;		
RX	MEDLINE=97262068; PubMed=9108482;		
RA	Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;		
RT	"Antigenic variation in Lyme disease borreliae by promiscuous		
RL	recombination of VMP-1like sequence cassettes.";		
DB	Cell 89:275-285(1997).		
DR	EMBL; U76406; AAC45178.1; -		
DI	InterPro; IPR000680; Borrelia_lipo.		
LF	Pfam; PF00921; Lipoprotein_2; 1.		
FT	NON_TER	1	
SQ	SEQUENCE	192 AA; 18519 MW; 78A5A2257F784AE8 CRC64;	

Query Match	63.28;	Score 148.5;	DB 2;	Length 192;
Best Local Similarity	69.68;	Pred. No. 2.2e-11;		
Matches 32; Conservative	6;	Mismatches 7;	Indels 1;	Gaps 1

QY 3 TNPIDAAIG-GSADRNAEAFDMMKKDDQIAAAMVLRGMAKDGQFAL 47  
 |||| |||| : : || : |||| |||| : |||| |||| : ||:  
 Db 138 TNPIDAAIGKGNENGAEFRDEMKDDQIAAAILRGMAKDGKFAY 183

RESULT 7  
007061

ID	007061	PRELIMINARY;	PRT;	217 AA.
AC	007061			
DT	01-JUL-1997	(TREMBLrel. 04, Created)		
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	VWP-LIKE SEQUENCE PROTEIN VLSE (FRAGMENT).			
OS	Vlse.			
OS	<i>Borrelia burgdorferi</i> (Lyme disease spirochete).			
OC	Plasmid lp28.1.			
OC	Bacteria; Spirochaetales; Spirochaetaceae; <i>Borrelia</i>			
NCBI_TaxID=139;				

SEQUENCE FROM N.A.			
RC STRAIN-B31-M4B4A;			
RX MEDLINE=9762068; PubMed=9108482;			
RA. Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;			
RT "Antigenic variation in Lyme disease borreliae by promiscuous			
RT recombination of VMP-1-like sequence cassettes.";			
RL Cell 89:275-285(1997).			
RL EMBL; U84562; AAC45199.1;			
DR InterPro: IPR000680; Borrelia_lipo.			
DR Pfam; PF00921; Lipoprotein_2; 1.			
KW plasmid.			
KM			
FT NON_TER	1	1	
FT NON_TER	217	217	
SQ SEQUENCE	217 AA;	20966 MW;	619E8CAB8B3A97449 CRC64;

Query Match	62.8%;	Score 147.5;	DB 2;	Length 217;
Best Local Similarity	73.3%;	Pred. No. 3.5e-11;		
Matches 33; Conservative	4;	Mismatches 7;	Indels 1;	Gaps 1

**Qy**      4 NPIDAIG-GSADRNAEAFDKMKDDQIAAAVLRGMAGDGQFAL 47  
         ||| |||| :|| | | |||||||: |||||:||:  
**Db**     157 NPIAAATGKGNADDGAIEFGDGMKKDDQIAAIALRGMAGDKGFAY 201

SEQUENCE	358 AA	36009 MW	6D22AE9NAA076EEC CRC64
SO	SEQUENCE	358 AA	36009 MW
KW	Plasmid		
DR	Pfam: PF00921: LipoProtein_2.1.		
DR	InterPro: IPR000680: Borrelia_lipo.		
DR	EMBL: U84556: AAC45764.1		
RL	Cell 89:275-285(1997).		
RT	"Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes."		
RA	Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.		
RX	MEDLINE:97262068; PubMed-9108482;		
RC	STRAIN-B31-MIE4;		
RP	SEQUENCE FROM N.A.		
RN	(1)		
OX	NCBI_TaxID=139;		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.		
CS	Plasmid lp28-1.		
GN	<i>Borrelia burgdorferi</i> (Lyme disease spirochete).		
DE	VMP-LIKE SEQUENCE PROTEIN VLSE.		
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)		
DT	01-JUL-1997 (TREMBLrel. 04, last sequence update)		
AC	007055;		
ID	007055	PRELIMINARY;	PRT; 358 AA.
RESULT	8		

Query Match	61.9%;	Score 145.5;	DB 2;	Length 358;
Best Local Similarity	69.6%;	Pred. No. 1.1e-10;		
Matches 32; Conservative	5;	Mismatches 8;	Indels 1;	Gaps 1

QY 3 TNPIDAAG-GSADRNAEAFDKMKDDQIAAAVYLRGMAKDGQFAL 47  
 ||| ||| : || ||||| : ||||| : ||  
 Db 253 TNPIDAAAGKGNENGAEEFGDMKKDDQIAAAIALRGMAKDGKFAV 298

RESULT	9		
006882		*	
ID	006882	PRELIMINARY;	PRT; 190 AA

DT 01-JUL-1997 (TREMBLERel. 04 Created)  
DT 01-JUL-1997 (TREMBLERel. 04, last sequence update)  
DE 01-JUN-2001 (TREMBLERel. 17, last annotation update)  
DE PURTATIVE VLS RECOMBINATION CASSETTE VLS5 (FRAGMENT).  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxId=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B31 (ATCC 35210) CLONE 5A3;  
RX MEDLINE-97262068; PubMed-9108482;  
RA Zhang J.R., Hardham J.M., Bahbout A.G., Norris S.J.;  
RT "Antigenic variation in Lyme disease borreliae by promiscuous  
RT recombination of VMP-like sequence cassettes.";  
RL Cell 89:275-285(1997).  
DR EMBL: U76406; AAC5179.1; "  
DR InterPro: IPR000680; Borrelia\_lipo.  
DR Pfam: PF00921; Lipoprotein\_2; 1.  
FT NON\_TER 1  
FT NON\_TER 190  
SQ SEQUENCE 190 AA; 18376 MW; 898EFF84308EAA7A CRC64;

Query Match	61.5%	Score 144.5;	DB 2;	Length 190;
Best Local Similarity	71.1%;	Pred. No. 7.2e-11;		
Matches	32;	Conservative	5;	Mismatches 7;
			Indels	1;
			Gaps	1

QY 4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGFAL 47  
 ||| ||| : || : | ||||| : ||||| : |||  
 Db 138 NPAAAIKGKNADDGADFGDGMKKDDQIAAAIALRGMAKDKGFAY 182

04573.C38





